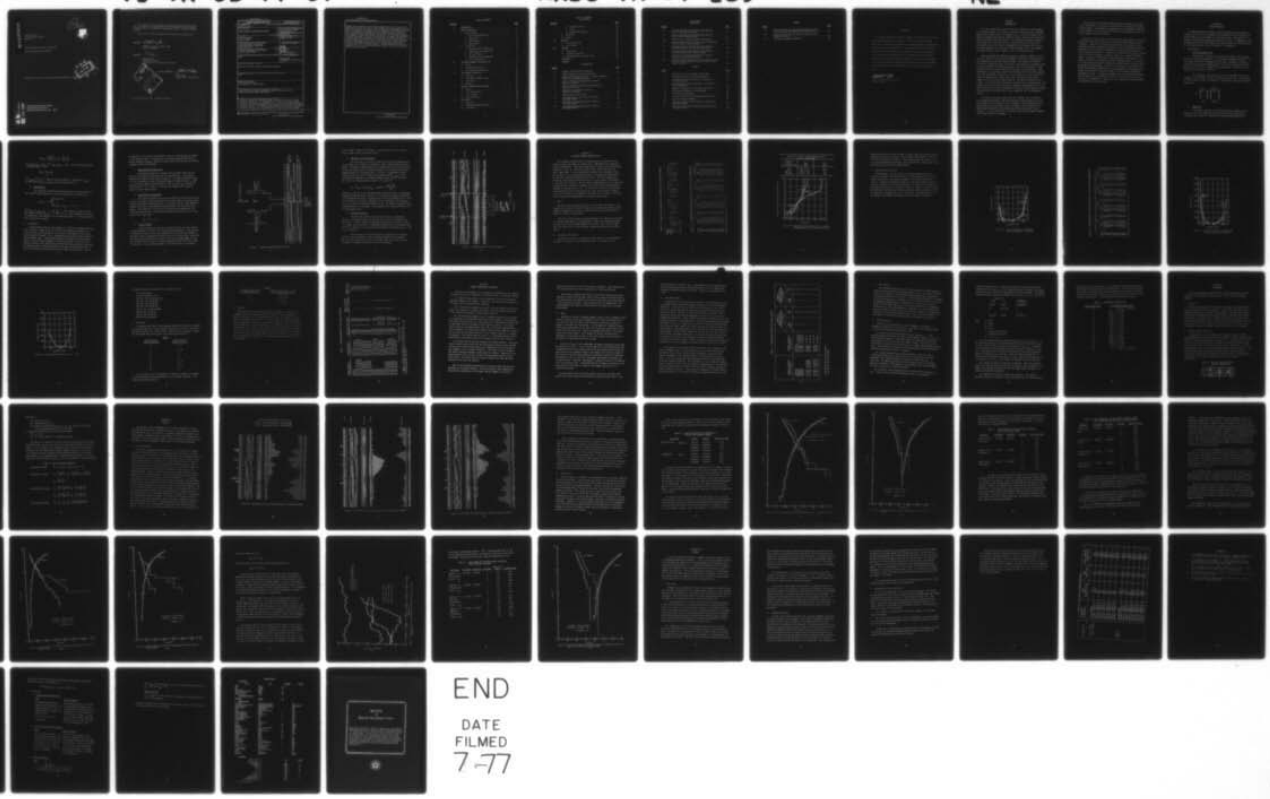
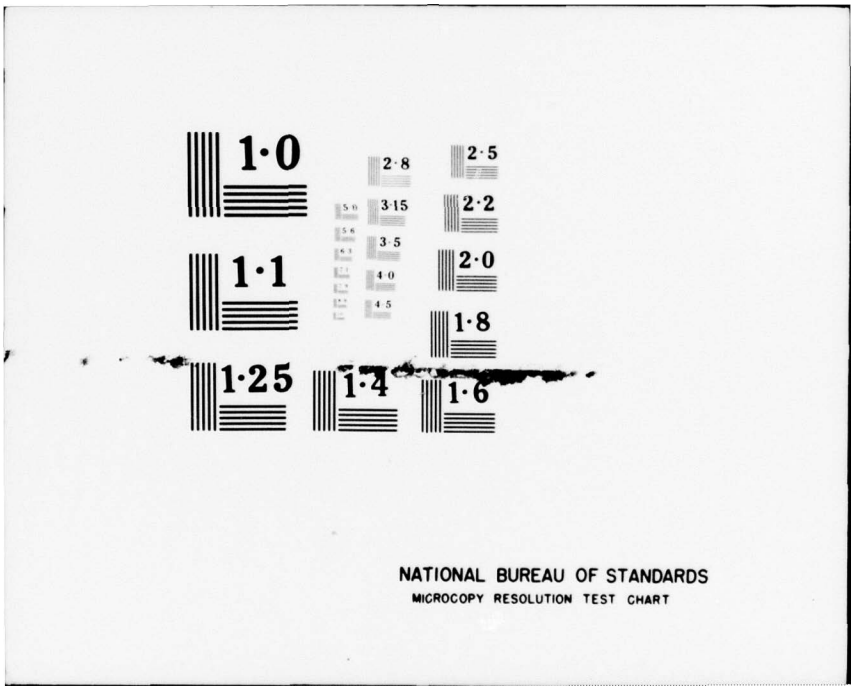
AD-A040 827 TEXAS INSTRUMENTS INC DALLAS F/G 9/4
REMOTE TERMINAL SPEAKER VERIFICATION.(U)
MAY 77 B G SECREST, R E HELMS F30602-76-C-0112
UNCLASSIFIED TI-TR-08-77-07 RADC-TR-77-169 NL

1 OF 1
ADA
040827



END

DATE
FILMED
7-77


1·0
1·1
1·25
1·4
1·6
1·8
2·0
2·2
2·5
2·8
3·15
3·5
4·0
4·5
5·0
5·6
6·3
NATIONAL BUREAU OF STANDARDS
MICROCOPY RESOLUTION TEST CHART

RADC-TR-77-169
Final Technical Report
May 1977

REMOTE TERMINAL SPEAKER VERIFICATION

Texas Instruments Incorporated

Approved for public release; distribution unlimited.

ROME AIR DEVELOPMENT CENTER
Air Force Systems Command
Griffiss Air Force Base, New York 13441

This report has been reviewed by the RADC Information Office (OI) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

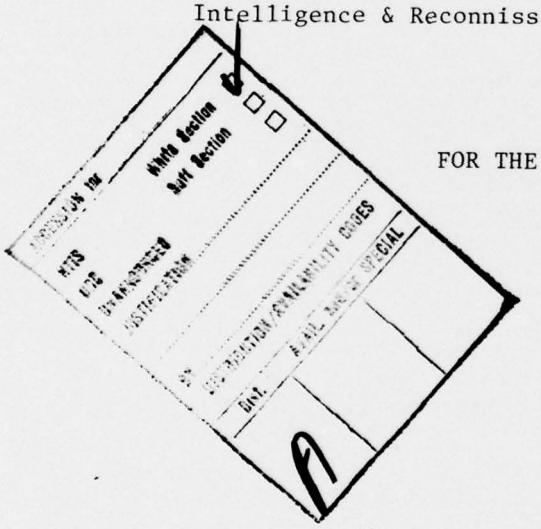
This report has been reviewed and is approved for publication.

APPROVED: Robert A. Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

APPROVED: H Davis

HOWARD DAVIS
Technical Director
Intelligence & Reconnissance Division



FOR THE COMMANDER: John P. Huss

JOHN P. HUSS
Acting Chief, Plans Office

Do not return this copy. Retain or destroy.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: RADC-TR-77-169	2. GOVT ACCESSION NO. / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE *(and Subtitle)*: REMOTE TERMINAL SPEAKER VERIFICATION.	5. TYPE OF REPORT & PERIOD COVERED: Final Technical Report, Jan 76 - Jan 77
	6. PERFORMING ORG. REPORT NUMBER: TR-08-77-07
7. AUTHOR(s): Bruce G. Secrest, Ramon E. Helms	8. CONTRACT OR GRANT NUMBER(s): F30602-76-C-0112
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas Instruments Incorporated, 13500 North Central Expressway, Dallas TX 75222	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 63714F, 681E0515
11. CONTROLLING OFFICE NAME AND ADDRESS: Rome Air Development Center (IRAP), Griffiss AFB NY 13441	12. REPORT DATE: May 1977
	13. NUMBER OF PAGES: 63
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*: Same	15. SECURITY CLASS. *(of this report)*: UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE: N/A

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

Same

18. SUPPLEMENTARY NOTES

RADC Project Engineer:
Captain Robert A. Curtis (IRAP)

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Speech Processing, Pattern Recognition, Speaker Verification, Voice Authentication, Personal Identification

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

A study was conducted to develop a speaker verification system for use over a degraded channel such as a telephone line. A test of the current speaker verification technology was performed on a set of data which had been processed through the RADC Digital Communications Experiment Facility (DICEF) to simulate a telephone channel. The simulated channel introduced an amplitude distortion, phase delay, and noise onto the analog data set. The noise did not present any particular problem other than to raise the spectral errors of both the true

→ next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

347650

speakers and the impostors, and the phase delay was also not a factor. The amplitude distortion was found to cause problems with time registration of the phrases and to result in an overall loss of information for speaker discrimination. To compensate the speaker verification system for these problems, a band-limited spectrum corresponding to the average band-pass characteristic of a telephone line was used for time registration, and the channel resistant pitch-period was added as an additional attribute for speaker discrimination. A limited experiment consisting of 16 speakers was conducted using the compensated speaker verification system. The results of the limited experiment were very encouraging in that a one percent true speaker rejection rate and a one percent impostor acceptance rate were obtained with a four phrase strategy. This exceeds the requirements of the Base Installation and Security System (BISS) of a one percent true speaker rejection rate and a two percent impostor acceptance rate.

TABLE OF CONTENTS

TABLE OF CONTENTS
(continued)

ILLUSTRATIONS

ILLUSTRATIONS
(continued)

TABLES

TABLES

EVALUATION

A study was conducted to develop a speaker verification system for use over degraded communication channels. A test of the current speaker verification technology was performed on a degraded channel and compensation techniques were then developed, i.e., a band-limited spectrum for time registration and the addition of the pitch-period error to the spectral error for speaker discrimination. This configuration was tested with a limited data set, and the BISS specifications of one percent true speaker rejection and two percent impostor acceptance were met with a four-phrase strategy employing the spectral error, plus the relative pitch-period error.

Robert A Curtis

ROBERT A. CURTIS, Captain, USAF
Project Engineer

SECTION I
INTRODUCTION

This report describes a study that has as its objective the development of an automatic speaker verification system that can be used over a degraded communication channel such as a telephone line. A typical telephone line is known to introduce band-limiting, phase distortion, and noise onto the signal being transmitted. The problem then is to compensate the method of speaker verification to minimize the effects of these distortions on the system. The methods used in this study to compensate for the channel conditions of a simulated telephone line resulted in a 99% acceptance rate of true speakers and a 99% rejection rate of impostors for a limited-test data set.

Texas Instruments has conducted research on automatic speaker verification for a number of years and has succeeded in moving the technology out of the laboratory and into a working environment. For example, an automatic entry control system has been in operation in the Corporate Information Center at Texas Instruments for the past two years, and a similar system is being evaluated as part of a base installation and security system (BISS). These systems are in a relatively noise-free environment and utilize good communication lines and dynamic microphones. However, there are various military and commercial applications, such as credit card validation, which require speaker verification from a remote location. The obvious solution to this problem is to use the readily available telephone line for communication. The purpose then of the current study was to determine the degradation in performance resulting from using the telephone line and to develop methods of improving performance to the point where it was comparable with the nondegraded channel.

For basically the same reasons that the telephone line is attractive as a communication channel, the ready availability and economy of the telephone headset make it attractive as a microphone. These advantages must be weighed against the additional distortion obtained from the carbon microphone whose characteristics can change according to the physical abuse involved. It is not unrealistic to consider replacing the carbon microphone with a better quality microphone.

The problems of the telephone headset were not addressed to any extent in this study except to perform a very limited test over several telephones with one speaker. The methods of compensation for the degraded channel used in this study were also restricted to techniques which would work with the current speaker verification system.

Section II of this report will describe the current speaker verification system being used and will describe the time registration and verification procedures. Section III will describe the results of a telephone survey conducted by the Bell Telephone Laboratories in which they characterized the telephone lines with respect to attenuation distortion, delay distortion, and the signal-to-noise ratio. Section IV will describe a preliminary evaluation of the current speaker verification system on a simulated telephone channel. Section V will describe the methods used for compensation in the speaker verification system when used over a degraded channel. This section will also discuss several other methods that could be used, along with their advantages and disadvantages. Section VI will describe an experiment conducted to determine the degradation of performance due to the communication channel and to determine the performance of the system when compensation techniques are used. A simulated telephone line will be characterized which was obtained by using the RADC-DICEF facility. Section VII will give the results of this experiment. Section VIII will summarize the study and offer recommendations for further study.

SECTION II
SPEECH PROCESSING

The speech processing strategy at Texas Instruments has led an evolutionary life from the first methods used, although they are all based on the relative spectrum of the speech as a function of time. This section will describe the processing strategy used during this study. A comparison of the processing used for the BISS speaker verification system with that given here for the remote terminal study is presented in the appendix.

A. Preprocessing

1. Filter Bank Definition

The spectrum is obtained by processing the speech signal through an analog filter bank preceded by a high frequency preemphasis network. The filter bank consists of 16 bandpass filters, each followed by a fullwave rectifier and a four-pole lowpass Bessel filter with a 3 dB cutoff at 30 Hz. Each of the 16 filters is sampled 100 times per second. The analog filter characteristics are given in Table 1.

For processing, the top three filters are averaged and filter 14 replaced by this average. Filters 15 and 16 are set to zero. The resulting 14 filter outputs at each time sample are represented by the spectrum amplitude vector:

$$\bar{A}_j = \begin{Bmatrix} a_{1j} \\ a_{2j} \\ \cdot \\ \cdot \\ \cdot \\ a_{14,j} \end{Bmatrix} = \begin{Bmatrix} a_1(t_j) \\ a_2(t_j) \\ \cdot \\ \cdot \\ \cdot \\ a_{14}(t_j) \end{Bmatrix} .$$

2. Regression

It has been found that by eliminating the gross aspects of the spectrum, such as the slope and curvature, more clearly defined format frequencies are obtained.[1] Therefore, the spectrum amplitude vector is

TABLE 1. CHARACTERISTICS OF 16-CHANNEL FILTER BANK[1]

Center Frequency (Hz)	Bandwidth (Hz, at -6 dB)
350	300
450	300
555	310
670	340
790	380
940	400
1120	400
1320	400
1550	400
1810	400
2100	400
2420	400
2800	400
3200	400
3800	800
5000	1600

regressed by the first three elements of an orthonormal basis set:

$$(\bar{A}_j)_R = \bar{A}_j - \sum_{k=0}^{2} c_{jk}\, \bar{F}_k$$

where

$$\bar{F}_k = \left\{ \begin{matrix} f_{1k} \\ . \\ . \\ . \\ f_{14,k} \end{matrix} \right\} \qquad k = 0,\ 1,\ 2$$

$$f_{i0} = 1$$

$$f_{i1} = -\sin\left[\frac{(i - 1/2)}{14}\,\pi\right]$$

$$f_{i2} = -\cos\left[\frac{(i - 1/2)}{14}\,\pi\right] \qquad i = 1,\ 2,\ \ldots,\ 14$$

and

$$c_{jk} = \frac{1}{14} \sum_{m=1}^{14} a_{mj}\, f_{mk} \quad .$$

3. Normalization

The regressed amplitude vector is then normalized by a modified postregression standard deviation for time t_j:

$$\sigma_j^{*} = \sigma_{post_j} + \frac{1}{8} \max(\sigma_{post_n}) : \quad n = j - 8,\ \ldots,\ j + 8$$

where

$$\sigma_{pre_j} = \frac{1}{13}\left(\sum_{m=1}^{14} a_{mj}\, a_{mj} - c_{j0}^2 \right)$$

$$\sigma_{post_j} = \frac{1}{11}\left(\sum_{m=1}^{14} a_{mj}\, a_{mj} - \sum_{k=0}^{2} c_{jk}^2 \right) .$$

If $\sigma_{post_j}/\sigma_{pre_j} < R_{min}$, then $\sigma_{post_j}/\sigma_{pre_j} = R_{min}$. The resulting normalized amplitude vector is then:

$$(A_j)_N = \frac{1}{\sigma_j^*} (A_j)_R .$$

The σ_{post} at time t_j is used as the energy measure. The energy (σ_{post_j}) and the coefficients c_{j1} and c_{j2} are also normalized by σ_j^*.

4. Quantization

The regressed and normalized amplitude vector is then quantized to one of eight levels according to a set of quantization thresholds $\{\phi_{iq}\}$:

$$(a_{ij})_Q = q \quad \text{IFF} \begin{cases} (a_{ij})_N \geq \phi_{iq} \\ (a_{ij})_N < \phi_{i,q+1} \end{cases} \quad \text{for } q = 0, 1, \ldots, 7$$

where $\phi_{iq} < \phi_{i,q+1}$; $\phi_{i0} = -\infty$; and $\phi_{i8} = \infty$. The $\{\phi_{iq}\}$ were chosen so that the probability that $(a_{ij})_Q = q$ is 1/8 for all q. The regression coefficients are quantized in the same manner as the amplitude vector. A linear quantization is used for the energy.

B. Processing

A detailed description of the processing is given in the Speaker Verification III report,[1] pages 13 and 14. Briefly, the processing strategy is to first accurately time register critical points of the spectrum or points of greatest spectral change. This is accomplished by defining scanning patterns from enrollment data. These scanning patterns are scanned across the input data to find an optimum sequence of registration points. This can be considered a zeroeth order time warping to time align various points of the input and reference data. Recognition patterns are then formatted between two or more time registration points which are then used for verification. This is

equivalent to a first-order time warping in that it helps account for changes in the length of words. A description of the scanning pattern definition, optimal sequence strategy, recognition pattern definition and the verification strategy will now be presented.

1. Scanning Pattern Definition

The scanning patterns are formed from the spectral data and are used to time align the input data with the reference data. The scanning patterns are formed from the scanning spectrum: the scanning pattern formed at time t_j consisting of (a) the average of spectral data at times t_{j-1} and t_{j-2}, (b) the average of spectral data at times t_{j+1} and t_{j+2}, and (c) the difference (b) - (a). Figure 1 illustrates the formation of a scanning pattern from the spectral data.

2. Scanning Error Computation

The scanning patterns are used to determine the location of the time registration points or reference points in the input speech on a phrase basis. At each time sample a scanning pattern is formatted from the input speech and compared with each of the reference scanning patterns for the words in the phrase. This comparison is done using the squared error between the k'th reference scanning pattern and the scanning pattern defined at the j'th time sample of the input data:

$$e_{kj} = \| x_j - r_k \| \quad .$$

3. Valley Finding

Using the scanning errors as a function of time, an error function is thus generated for each of the reference scanning patterns. Each function is monitored for dips of sufficient magnitude to be considered as potential locations of the corresponding reference points in the input data. These dips are called valley points when the ratio of the local maximum (peak) and the dip is greater than or equal to a specified peak-to-valley (PV) ratio, which in this case is 1.3, and the magnitude of the valley point is less than or

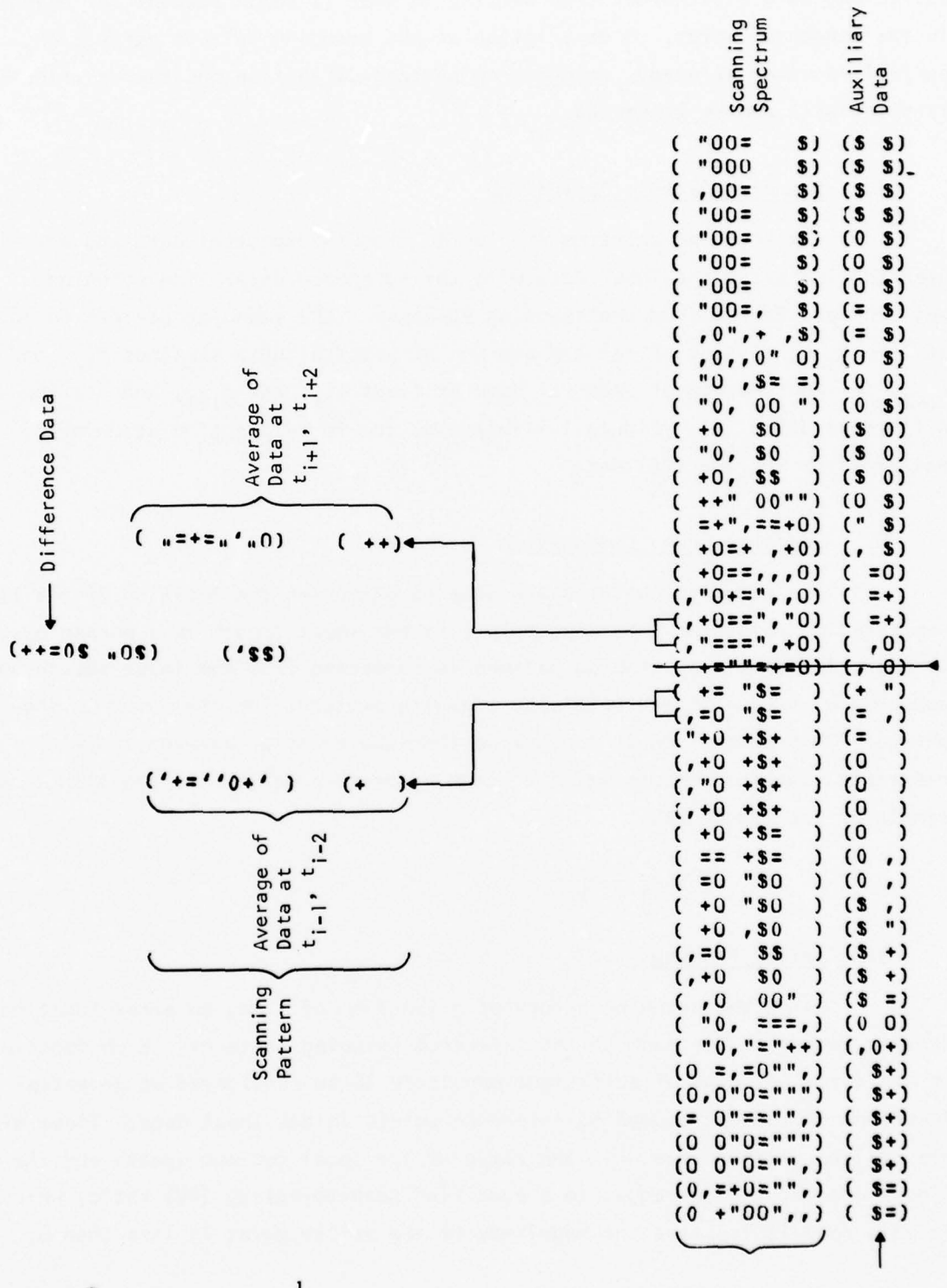


Figure 1.[1] Example Scanning Pattern Formation

equal to 200. A peak of (1/PV-ratio) * (valley point error) is required before another valley point can be found.

4. Reference Point Sequencing

The next step is to pick a valley point from each reference point in the phrase to fit together a sequence of eight reference points for the phrase. A sequence is determined for all combinations of valley points which satisfy the phrase specific minimum and maximum time distance restrictions between each pair of reference points. An error is determined for each reference point pair that is based on the scanning errors, e_{rp_k}, of the points and the expected time distance between the two points. The point pair error for reference points k and k + 1 is given by:

$$E_w = (e_{rp_k} + e_{min})(e_{rp_{k+1}} + e_{min})\left(1 + \beta \frac{\Delta t_k - \Delta \hat{t}_k}{\Delta \hat{t}_k}\right)$$

where $e_{min} = 40$, Δt_k is the distance between the points, $\Delta \hat{t}_k$ is the expected distance between the points, and $\beta (= 1)$ is a penalty assigned for deviations from the expected distance. These point pair errors are summed for all reference point pairs to obtain a sequence error for the phrase. This sequence error is limited by a maximum threshold of 100. The optimal sequence of eight reference points for the phrase is the combination of valley points with the minimum sequence error.

5. Verification Error

To obtain the verification error for a phrase, a recognition pattern is formatted between the time registration points of each word in the phrase. A sample recognition pattern is illustrated in Figure 2, where it is shown that the pattern is interpolated between the two reference points, which is in contrast to the BISS verification where the scanning error is used for verification.

The resulting verification patterns are compared with reference patterns, and a squared-error sum is obtained between the two patterns. This is called the spectral error for the word; when the four word errors are added, it becomes the spectral error for the phrase.

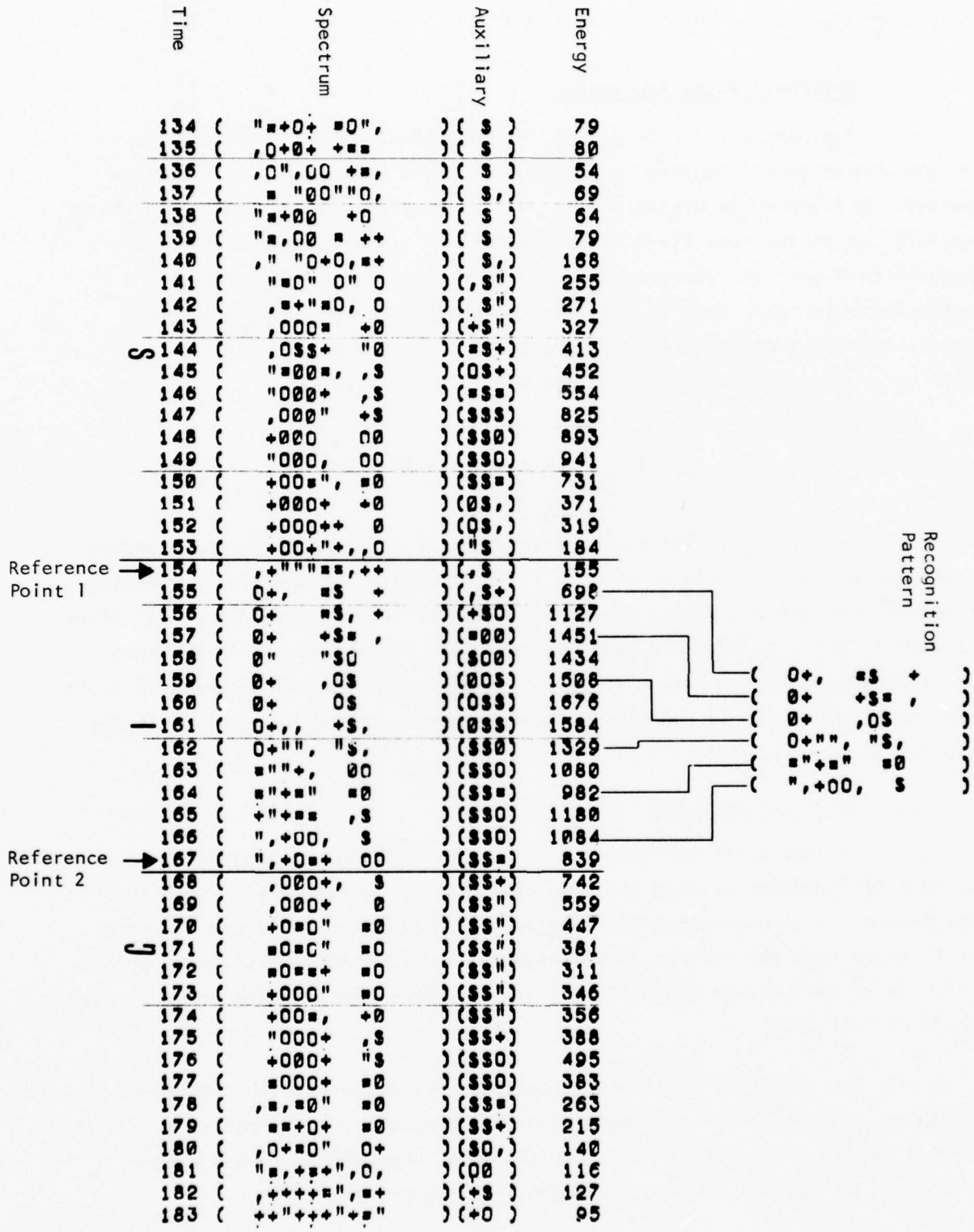


Figure 2. Example Recognition Pattern Formation

SECTION III

TELEPHONE CHANNEL CHARACTERISTICS

Bell Telephone Laboratories has conducted several system-wide transmission surveys since 1959, the latest being a 1969-1970 survey conducted by Duffy and Thatcher.[2] The survey separated the results into three mileage categories: short (0 - 180 miles), medium (180 - 725 miles), and long (725 - 2900 miles). Of the transmission characteristics discussed in the survey; noise, loss, attenuation distortion, and envelope delay distortion are the most important to speaker verification. The noise, attenuation distortion, and envelope delay distortion will be discussed in more detail in the next three sections. The loss can be compensated by using an amplifier and should cause no problem to speaker verification processing provided the response of the amplifier is flat. In a series of trials in the laboratory it was found that the loss was greater over the lines that went through a "dial 9" or outside line as opposed to an inside line or Centrex. The level of the signal can be compensated by an automatic gain control.

A. Noise

The noise study was broken into circuit noise and impulse noise. The impulse noise results are given in Table 2 and represent the number of noise impulses exceeding four different voltage levels received when a 2750 Hz signal at -12 dBm is transmitted.

The circuit noise results are given graphically in Figure 3 as the ratio of received signal power to C-notched noise. Note that the average signal-to-noise ratio for all mileage categories is 41 dB. Also, Figure 3 shows that less than 1% of all the lines have a signal-to-noise ratio less than 20 dB.

B. Attenuation Distortion

Attenuation distortion is a measure of the variation in loss caused by a change in frequency of the transmitted signal. The results of the

TABLE 2.[2] IMPULSE NOISE COUNTS ON TOLL CONNECTIONS (15-MINUTE INTERVAL)

Signal-to-Impulse-Noise Counter Threshold	All			0–180 Miles			180–725 Miles			725–2900 Miles		
	Mean	Med.*	S.D.	Mean	Med.	S.D.	Mean	Med.	S.D.	Mean	Med.	S.D.
+5	39 ± 21	4	128	32 ± 15	3	105	48 ± 44	7	150	74 ± 48	15	219
+1	18 ± 11	2	68	13 ± 6	1	40	27 ± 30	2	90	44 ± 32	7	148
−3	11 ± 8	1	56	7 ± 5	0	23	24 ± 26	1	115	24 ± 20	2	87
−7	7 ± 6	0	34	5 ± 4	0	19	15 ± 18	0	72	10 ± 11	0	28

* Med. indicates median

TABLE 3.[2] ATTENUATION DISTORTION RELATIVE TO 1000 Hz ON TOLL CONNECTIONS

Frequency (Hz)	All		0–180 Miles		180–725 Miles		725–2900 Miles	
	Mean (dB)	S.D. (dB)	Mean (dB)	S.D. (dB)	Mean (dB)	S.D. (dB)	Mean (dB)	S.D. (dB)
200*	11.8 ± 1.9	5.1	11.4 ± 2.4	5.1	13.7 ± 1.5	4.5	12.4 ± 2.4	5.0
250	6.6 ± 0.9	3.0	6.4 ± 1.1	2.7	8.0 ± 1.3	3.7	6.8 ± 1.6	3.1
300	4.1 ± 0.6	2.1	4.0 ± 0.7	1.9	4.8 ± 0.9	2.8	4.0 ± 0.9	2.1
400	2.3 ± 0.3	1.6	2.2 ± 0.4	1.4	2.8 ± 0.6	2.2	2.0 ± 0.3	1.4
600	1.1 ± 0.1	1.1	0.9 ± 0.1	0.9	1.6 ± 0.4	1.9	1.2 ± 0.2	0.8
800	0.5 ± 0.1	0.5	0.4 ± 0.0	0.5	0.7 ± 0.2	0.5	0.5 ± 0.1	0.4
1200	−0.1 ± 0.1	0.3	−0.1 ± 0.1	0.3	−0.3 ± 0.1	0.4	−0.3 ± 0.0	0.4
1400	−0.1 ± 0.1	0.6	0.0 ± 0.1	0.6	−0.3 ± 0.1	0.6	−0.3 ± 0.1	0.5
1700	0.3 ± 0.1	0.9	0.3 ± 0.1	0.9	0.1 ± 0.2	0.8	0.2 ± 0.2	0.8
2000	0.8 ± 0.1	1.1	0.8 ± 0.1	1.1	0.8 ± 0.2	1.1	0.7 ± 0.3	1.0
2300	1.4 ± 0.2	1.4	1.4 ± 0.3	1.3	1.4 ± 0.3	1.4	1.7 ± 0.5	1.4
2450	1.9 ± 0.4	1.6	1.8 ± 0.4	1.5	2.0 ± 0.4	1.6	2.4 ± 0.6	1.7
2750	3.7 ± 0.8	2.5	3.5 ± 1.0	2.5	4.1 ± 0.5	2.2	4.7 ± 1.0	2.3
2850	4.7 ± 1.1	3.0	4.4 ± 1.2	3.0	5.4 ± 0.6	2.6	6.1 ± 1.1	2.7
3000	6.9 ± 1.5	4.1	6.4 ± 1.6	4.1	8.1 ± 1.0	3.6	9.2 ± 1.9	4.3
3100	9.5 ± 1.7	5.7	9.0 ± 1.9	5.9	10.6 ± 1.5	4.7	11.6 ± 2.7	5.2
3200*	13.4 ± 2.1	7.8	12.9 ± 2.5	8.0	14.7 ± 2.1	6.8	15.2 ± 3.4	6.6
3300*	18.2 ± 2.7	9.5	17.6 ± 3.2	10.0	20.0 ± 3.1	8.0	19.8 ± 3.7	7.6
3400*	22.1 ± 3.0	9.2	21.2 ± 3.6	9.8	24.4 ± 2.8	6.4	25.1 ± 1.7	6.1

* Distortion values at these frequencies are at least as great as shown.

Results of Signal-to-C-Notched-Noise Ratio for Toll Connections

Connection Length (Airline Miles)	Mean (dB)	S.D. (dB)
All	40.6 ± 3.0	11.8
0-180	42.1 ± 3.8	13.0
180-725	36.5 ± 1.3	5.3
725-2900	35.4 ± 0.9	3.8

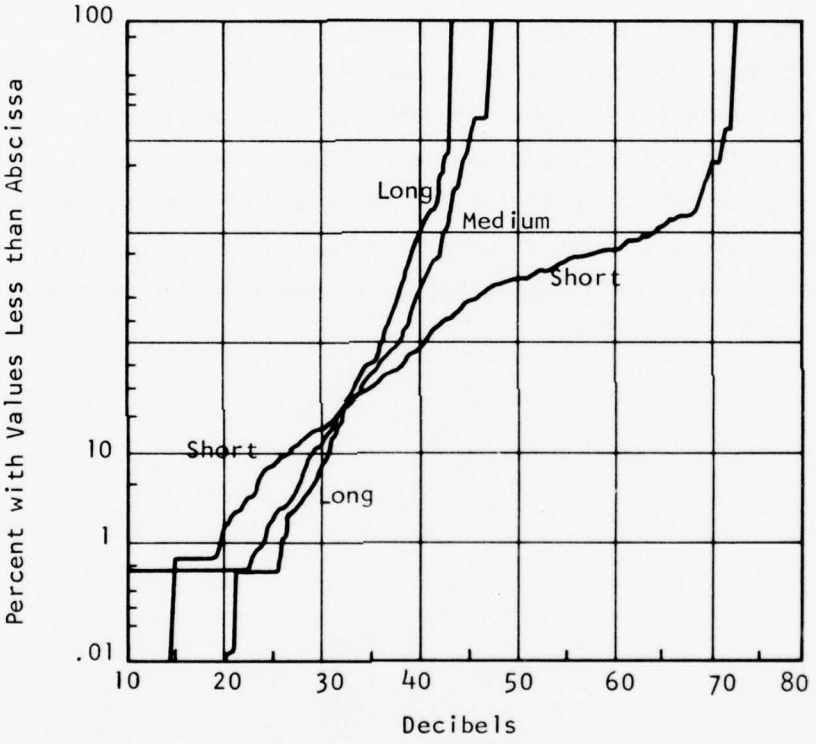


Figure 3.[3] Cumulative Distribution Curves: Signal-to-Notched Noise with C-Message Weighting

attenuation distortion study are shown in Table 3 where the mean and standard deviation are given as a function of the transmitting frequency. The locus of the mean is plotted in Figure 4. The average attenuation distortion is less than 3 dB between approximately 500 Hz and 2700 Hz. Beyond these ranges, the attenuation distortion increased rapidly.

C. Envelope Delay Distortion

Envelope delay is the derivative of the phase characteristic with respect to frequency, and the distortion is the envelope delay minus a constant delay term for a particular frequency, in this case the delay at 1700 Hz. The mean and standard deviation of the envelope delay distortion are given in Table 4 for various frequencies. The locus of the means of the envelope delay distortion is plotted in Figure 5. Again it can be seen that the distortion increases rapidly at the lower and higher frequencies. Since the spectrum is averaged over a 10 ms time period, the phase distortion is not thought to be especially damaging to the verification procedures.

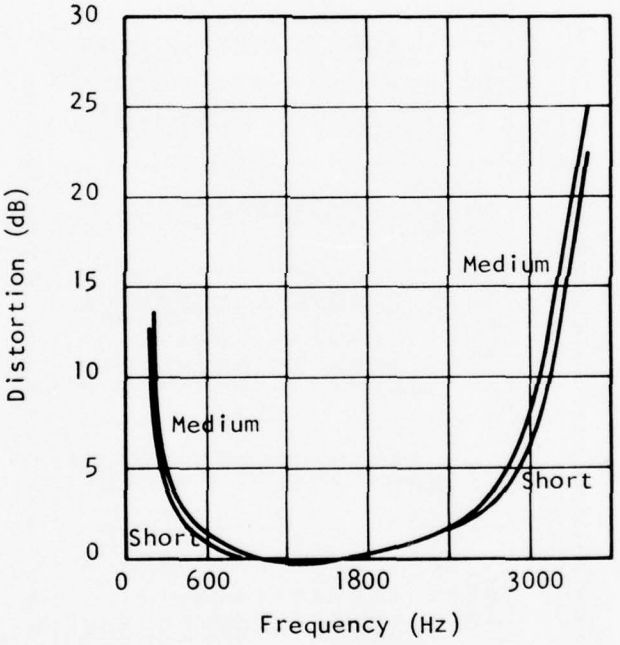


Figure 4.[2] Locus of Means for Attenuation Distortion Relative to 1000 Hz

TABLE 4.[2] ENVELOPE DELAY DISTORTION RELATIVE TO 1700 Hz ON TOLL CONNECTIONS

TABLE IX—ENVELOPE DELAY DISTORTION RELATIVE TO 1700-HZ ON TOLL CONNECTIONS

Frequency (Hz)	All		0–180 Miles		180–725 Miles		725–2900 Miles	
	Mean (μsec.)	S.D. (μsec.)	Mean (μsec.)	S.D. (μsec.)	Mean (μsec.)	S.D. (μsec.)	Mean (μsec.)	S.D. (μsec.)
200*	5187 ± 566	2672	4580 ± 518	2461	7526 ± 404	1851	7505 ± 473	2422
250*	3934 ± 410	2010	3384 ± 326	1727	5866 ± 417	1595	5880 ± 314	1870
300	3290 ± 289	1660	2816 ± 209	1407	4884 ± 384	1375	4901 ± 297	1510
400	2091 ± 221	1220	1695 ± 128	930	3413 ± 341	1215	3163 ± 218	1144
600	843 ± 96	583	656 ± 43	430	1467 ± 183	628	1335 ± 127	592
800	392 ± 50	342	290 ± 20	263	737 ± 114	371	649 ± 88	350
1000	190 ± 28	206	133 ± 15	165	380 ± 73	227	335 ± 53	209
1200	80 ± 16	125	48 ± 10	103	187 ± 45	139	156 ± 32	128
1400	17 ± 5	74	3 ± 8	66	63 ± 16	83	56 ± 15	76
2000	51 ± 20	67	50 ± 18	62	36 ± 26	66	80 ± 51	95
2300	175 ± 47	136	152 ± 43	122	226 ± 54	133	273 ± 76	180
2450	284 ± 65	179	248 ± 64	159	363 ± 48	153	442 ± 89	230
2750	577 ± 120	339	485 ± 102	276	811 ± 99	273	934 ± 189	457
2850	729 ± 144	420	616 ± 120	338	1017 ± 137	348	1166 ± 263	573
3000	1041 ± 183	570	889 ± 164	456	1437 ± 144	468	1614 ± 303	816
3100	1335 ± 241	728	1128 ± 217	578	1903 ± 153	585	2071 ± 329	993
3200	1636 ± 330	956	1319 ± 279	697	2475 ± 191	750	2734 ± 414	1285
3300*	1919 ± 461	1227	1526 ± 363	917	3208 ± 343	1095	3333 ± 395	1356
3400*	2367 ± 693	1645	1935 ± 556	1277	4040 ± 553	1634	4248 ± 752	2018

* A significant percentage of connections were not measurable at these frequencies.

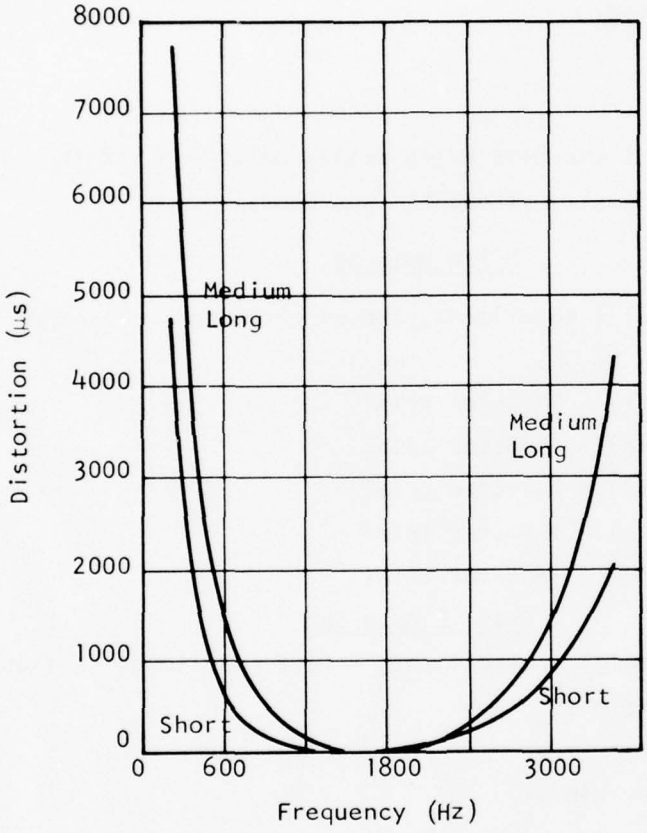


Figure 5. Locus of Means for Envelope Delay Distortion Relative to 1700 Hz

SECTION IV
PRELIMINARY EVALUATION

As the basis for evaluation of the effects of a telephone channel on the current speaker verification technology, a very limited experiment was conducted. The test data was run through the RADC-DICEF facility to introduce noise and channel coloration typical of a telephone channel and the performance was evaluated.

A. Data Set

A limited subset of the BISS Mitre analog data base and the BISS Phase I data set collected at Texas Instruments[3] was used.

Mitre Data Set

(1) No. 432 (male) - enrollment; 2 post enrollment sessions; 1 execution session
(2) No. 5088 (male) - impostor trial
(3) No. 3387 (male) - impostor trial
(4) No. 1368 (male) - impostor trial
(5) No. 9062 (male) - impostor trial
(6) No. 3263 (male) - impostor trial

Phase I Data Set

(7) Dan Daniel (male) - enrollment; 4 post enrollment sessions; 4 execution sessions.

B. Simulated Telephone Lines

The above analog data set was processed through the RADC-DICEF facility to introduce noise, a flat line, and a 4A line coloration. The flat line had a virtually flat response. The amplitude distortion of the 4A line or simulated telephone line is shown in Figure 6. The envelope delay distortion is shown in Figure 7. As can be seen by comparing these figures with Figures 4 and 5, they simulate the average telephone line reasonably well.

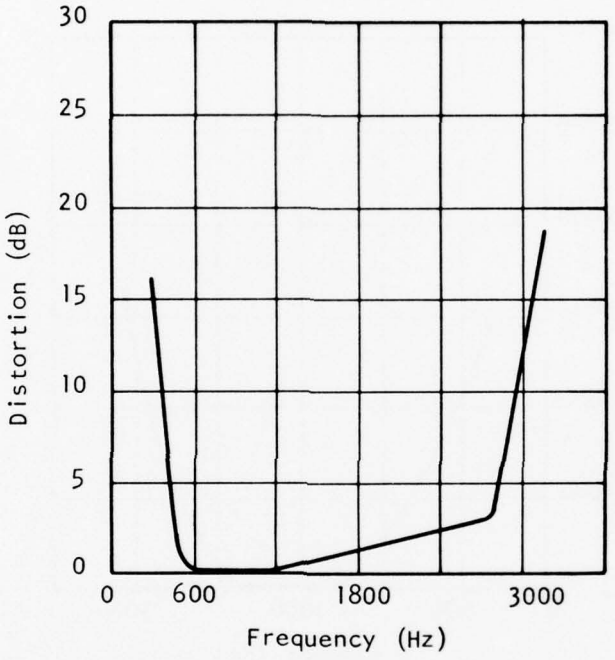


Figure 6. Attenuation Distortion for 4A Line Relative to 1000 Hz

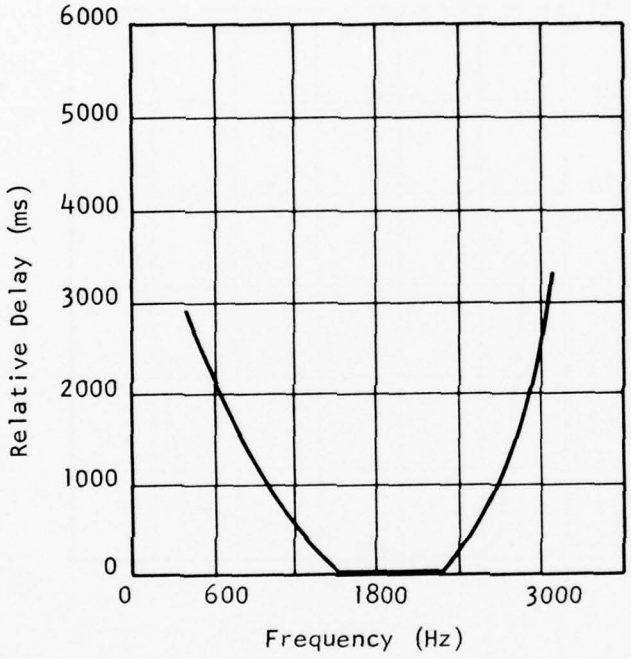


Figure 7. Relative Delay for 4A Line for $\tau = 0.0$

The channel conditions considered for the experiment are:

(1) Original tape
(2) 4 kHz lowpass filter
(3) Flat line; no noise added
(4) Flat line; 10 dB S/N
(5) Flat line; 20 dB S/N
(6) Flat line; 30 dB S/N
(7) 4A line; no noise added
(8) 4A line; 10 dB S/N
(9) 4A line; 20 dB S/N
(10) 4A line; 30 dB S/N

C. Experiment

To evaluate the effect of the degraded channel conditions on the speaker verification algorithm, various channel conditions were used for enrollment and then execution. For Type I errors, (rejection of true speaker), speaker No. 432 and Dan Daniel were used in the following combinations:

Type I

Enrolled with Channel Condition	Execution with Channel Condition
1	1 - 10
2	2
3	1 - 10
4	4
5	5
6	6
7	1 - 10
8	8
9	1 - 10
10	10

For the Type II errors (acceptance of imposters), speakers 2 through 6 of the Mitre data set were used against the true speaker (No. 432). The channel conditions used were:

	Type II
(No. 432) Enrolled with Channel Condition	Imposter Trials of 2 - 6 with 1 for Channel Conditions
1	1, 3, 7, 9
3	1, 3, 7, 9
7	1, 3, 7, 9
9	1, 3, 7, 9

D. Results

The results of the experiment are shown in Table 5. A spectral error threshold of 200 was chosen and Type I and Type II errors calculated. A Type I error occurs when the spectral error for the true speaker is greater than the threshold, and a Type II error occurs when the impostor spectral error is less than or equal to the threshold. Because of the limited nature of this experiment, the results have very little meaning except in a qualitative sense. For example, Table 5 shows that great difficulty was encountered in registering reference points for the true speakers when the execution was on a different channel condition than the enrollment. In the next section compensation methods will be discussed with which these phrases can be registered.

TABLE 5. SPEAKER PERFORMANCE IN PRELIMINARY EXPERIMENT

Channel Condition for Enrollment	Channel Condition for Enrollment	Number of Type I Errors* Dan Daniel	Number of Type I Errors* No. 432	Number of Phrases not Registered Daniel	Number of Phrases not Registered No. 432	Number of** Type II Errors
Original	Original	0 (0%)	0 (0%)			0 (0%)
Original	Flat Line; No Noise	10 (32%)	1 (11%)	-	1	0 (0%)
Flat Line; No Noise	Flat Line; No Noise	1 (3%)	0 (0%)	-	-	1 (2%)
Flat Line; 10 dB S/N	Flat Line; 10 dB S/N	6 (19%)	0 (0%)	-	-	0 (0%)
Flat Line: 20 dB S/N	Flat Line; 20 dB S/N	1 (3%)	0 (0%)	-	-	1 (2%)
Flat Line; 30 dB S/N	Flat Line; 30 dB S/N	1 (3%)	0 (0%)	-	-	0 (0%)
4A Line; No Noise	4A Line; No Noise	0 (0%)	0 (0%)	-	-	1 (2%)
4A Line; No Noise	4A Line; 20 dB S/N	0 (0%)	0 (0%)	-	-	0 (0%)
4A Line; 10 dB S/N	4A Line; 10 dB S/N	1 (3%)	0 (0%)	-	-	1 (2%)
4A Line; 20 dB S/N	4A Line; 20 dB S/N	0 (0%)	0 (0%)	-	-	2 (4%)
4A Line; 20 dB S/N	4A Line; No Noise	0 (0%)	0 (0%)	-	-	1 (2%)
4A Line; 30 dB S/N	4A Line; 30 dB S/N	0 (0%)	0 (0%)	-	-	2 (4%)
Original	4 kHz LP Filter	0 (0%)	-	-	-	-
Original	Flat Line; 10 dB S/N	19 (61%)	2 (22%)	2	1	-
Original	Flat Line; 20 dB S/N	9 (29%)	0 (0%)	-	-	-
Original	Flat Line; 30 dB S/N	10 (32%)	1 (11%)	-	1	-
Original	4A Line; No Noise	13 (42%)	9*** (100%)	8	9	-
Original	4A Line; 10 dB S/N	24 (77%)	9*** (100%)	20	9	-
Original	4A Line; 20 dB S/N	17 (55%)	9*** (100%)	15	9	-
Original	4A Line; 30 dB S/N	12 (39%)	9*** (100%)	5	9	-
4A Line; No Noise	Original	4 (13%)	9*** (100%)	2	9	-
4A Line; No Noise	4 kHz LP Filter	3 (10%)	-	2	-	-
4A Line; No Noise	Flat Line; No Noise	8 (26%)	6 (67%)	1	4	-
4A Line; No Noise	Flat Line; 10 dB S/N	23 (74%)	9*** (100%)	23	9	-
4A Line; No Noise	Flat Line; 20 dB S/N	11 (35%)	6 (67%)	4	4	-
4A Line; No Noise	Flat Line; 30 dB S/N	7 (23%)	6 (67%)	2	4	-
4A Line; No Noise	4A Line; 10 dB S/N	2 (6%)	0 (0%)	-	-	-
4A Line; No Noise	4A Line; 30 dB S/N	0 (0%)	0 (0%)	-	-	-

Threshold Spectral Error = 200

* Number of phrases with spectral error greater than 200
** Number of phrases with spectral error less than or equal to 200
*** No sequences were registered

SECTION V
CHANNEL COMPENSATION TECHNIQUES

A variety of philosophies can be employed to compensate for the channel coloration and noise introduced by a telephone channel. These would include:

(1) Replacing the channel with a twisted pair so as to eliminate the degrading effects of the channel. Obviously, this would not be practical in all cases, such as between distant locations.

(2) Eliminating the channel by sending a digital signal over the telephone line instead of the analog voice data. This would involve placing a filter bank and preprocessor at each sending location.

(3) Calibrating the channel so as to weight various parts of the spectrum according to the amplitude distortion of the telephone line being used. This would involve some sort of signal generator at each sending location and an additional piece of equipment at the receiving location. An ultimate loss of spectral information would result due to the severe amplitude distortion outside of the approximate spectral range of 500 to 2700 Hz. Another way to calibrate the channel would be to use a channel equalizer to flatten the spectrum. This method usually results in a considerable amount of noise being added to the channel outside the frequency range of 500 to 2700 Hz.

(4) Normalizing the data and expanding the speech measures. This might involve normalizing the data with the noise energy during quiescent periods and also using a band-limited spectrum corresponding to the narrow band of the telephone line for speaker processing. This would involve a change in the speech processing software at the receiving location. Expansion of the speech measures might include a channel resistant speech measure such as pitch period. This would involve a piece of hardware and some additional software at the receiving location.

Most of the procedures mentioned would involve an additional piece of equipment at the sending location. This may be good or bad, depending on the particular application of the remote terminal speaker verification

system and the ultimate cost of the piece of equipment. Some combination of the above procedures may also be appropriate in some applications.

The method of compensation chosen in the present study was the latter approach of normalizing the data, that of using a band-limited spectrum for time registration, and adding the pitch period for additional discrimination in the verification process. These measures allow the compensation to be accomplished at the receiving location and will be amplified in the next few sections.

A. Noise

According to the Bell Telephone Company survey of their telephone lines,[2] two major types of noise can be expected; circuit noise and impulse noise. It is not felt that the impulse noise problem is particularly severe to speaker verification for two reasons: (1) the spectrum is obtained from a filter bank sampled every 10 ms, each filter being the average over a 10 ms interval of the time series. One or two impulses would have little effect on the filter bank output. (2) A sequential decision strategy can be employed for verification so that if the noise affects the verification on one phrase, an additional phrase can be used.

The circuit noise is more bothersome and several techniques can be used to overcome its effects. First, through regression and normalization by the standard deviation, the constant noise can be eliminated and the effects of white noise can be reduced as shown in the speaker verification II report.[3] The results of the preliminary evaluation discussed in Section IV show that the spectral errors for both the true speaker and impostors increase with increasing noise. Therefore, some method is needed to normalize the spectral error for the noise level. One method would be to measure the energy in the input signal during times of silence by the speaker and use this in a normalization scheme.

Another method of minimizing the effects of noise is to measure the energy in the input signal during periods of silence and then filter it out

of the signal with a Wiener filter. Unfortunately, the lab speech system does not allow processing the signal between the time it is sampled and processed through the analog filter bank. Therefore, this method was not tried.

B. Time Registration

As can be seen by the preliminary evaluation, time registration of the true speakers was a significant problem, particularly on the cross-channel conditions. Since the flat lines with noise added seemed to give little problem, it became apparent that the amplitude distortion was preventing time registration. Therefore, a band-limited spectrum was used. From Figure 6 it can be seen that the amplitude of the 4A line is fairly flat between 500 and 2750 Hz. Table 1 shows the frequency characteristics of the 16 analog filters. As mentioned earlier, the top three channels are averaged and used for filter 14. Therefore, the top four filters and the bottom two filters were not used, leaving 10 filters covering the frequency range from 400 to 2620 Hz. New quantization limits were obtained, and the spectrum was renormalized. The preliminary evaluation test was rerun for the difficult time registration channel conditions. All but one of the phrases that were not previously registered were registered using the 10-channel spectrum. The results are shown in Table 6, where the number of phrases not registering a sequence for the true speakers are shown using 14 filters and 10 filters.

The frequency range of 400 to 2620 Hz seems also to be appropriate for actual telephone lines. This was borne out in a very small survey using Dan Daniel (a participant in the original BISS Phase I experiment) over several telephones in the laboratory where inside and outside lines were used. All of his phrases over these telephone lines were time-registered using the 10-channel spectrum. This experiment will be discussed in greater detail in Section VIII.B. That this method of compensating for the amplitude distortion should also work rather well on a typical telephone line can be seen by looking at Table 3. In the range of frequencies from 400 to 2620 Hz the mean distortion is less than 2.6 dB with a standard deviation of 1.6 dB.

TABLE 6. MISREGISTERED PHRASES FOR TRUE SPEAKERS

CHANNEL CONDITION FOR ENROLLMENT	CHANNEL CONDITION FOR EXECUTION	NUMBER OF PHRASES WITH NO SEQUENCE REGISTERED - 14 FILTERS		NUMBER PHRASES WITH NO SEQUENCE REGISTERED - 10 FILTERS	
		DAN DANIEL*	NO. 432†	DAN DANIEL	NO. 432
ORIGINAL	ORIGINAL	0	0	0	0
ORIGINAL	FLAT LINE - NO NOISE	-	1	-	0
ORIGINAL	4A LINE - NO NOISE	8	9	1	0
ORIGINAL	4A LINE - 10 DB S/N	-	9	-	0
ORIGINAL	4A LINE - 20 DB S/N	15	9	0	0
4A LINE - NO NOISE	4A LINE - NO NOISE	-	0	-	0
4A LINE - NO NOISE	ORIGINAL	-	9	-	0
4A LINE - NO NOISE	4A LINE - 20 DB S/N	-	0	-	0
4A LINE - 20 DB S/N	4A LINE - 20 DB S/N	-	0	-	0
4A LINE - 20 DB S/N	ORIGINAL	-	9	-	0
4A LINE - 20 DB S/N	4A LINE - NO NOISE	-	0	-	0

* Dan Daniel has 31 total phrases

† No. 432 has 9 total phrases

C. Verification

The 10-channel spectrum is used in forming the verification patterns in addition to being used for time registration. Thus, a net loss of information for verification is suffered when comparing with the 14 channels of information available on a flat line. To bring the performance of the speaker verification system over a telephone line up to that obtained on a flat line, an additional speech measure is needed for the verification. Several authors[4,5] have shown that pitch is relatively insensitive to the bandpass characteristics of a telephone channel, although it is more susceptible to mimics than other measures, such as the spectrum. Therefore, a combination of spectral error and pitch period error is used for the verification. A description of the pitch extraction method is discussed in Section V.D.

D. Pitch Extraction

The pitch is extracted only in certain segments of a phrase, i.e., between the marked vowels of each word in the phrase. The following steps are used in extracting an optimal pitch track:

(1) Obtain pitch period estimates at each time step, t_i. Pitch period estimates are made every 10.05 ms (the sampling interval of the filter bank) using the Cepstrum. A 201-point (30.15 ms) data block consisting of the preceding, current, and next time frames is used. The points of this data block are obtained from an A/D filter with a 6667 samples/second sampling rate. The data block is low-pass filtered and passed through a Hamming window before a 256-point Cepstrum is computed.

(2) Compute candidate pitch period estimates at each time step, t_i. A peak-finding algorithm is used to find all of the peaks of sufficient magnitude of the Cepstrum between 52 and 202 Hz (4.95 to 19.2 ms). In particular, the maximum peak plus all others greater than 0.25 X (maximum peak) are considered. These peaks are the candidate pitch period estimates at time t_i, along with an unvoiced estimate.

(3) Calculate transition penalty for pitch period from time step t_{i-1} to t_i. Each of the pitch-period estimates at time t_i are assessed a

transition penalty for the transition from each of the saved pitch-period estimates at time step t_{i-1}. This transition penalty is based on the smoothness of the pitch track and the magnitude of the Cepstral peaks. The penalty assessed to each pitch-period estimate from time step t_{i-1} to t_i is given by:

Transition t_{i-1}	Transition t_i	Penalty
Voiced	Voiced	$\frac{1 + \beta(\Delta\tau/\tau)^2}{\rho}$
NA	Unvoiced	k_1
Unvoiced	Voiced	$\frac{1 + \beta\tau_1^2}{\rho} + k_2$

where $\beta = 200$

$k_1 = 0.08$

$k_2 = 0.19$

$\tau_1 = 0.0$

τ = pitch-period estimate

ρ = Cepstral peak value

$\Delta\tau = \tau_i - \tau_{i-1}$

(4) Calculate cumulative pitch-period penalty up to time step t_i. As the transition penalty is calculated for each pitch period estimate at time t_i with all of the saved pitch-period estimates at time t_{i-1}, it is added to the cumulative error associated with the pitch-period estimates at time t_{i-1}. The eight current pitch-period estimates with the lowest cumulative penalty are saved at this time step. These new cumulative penalties are saved with the pitch-period estimates at time t_i along with back-pointers to the estimates at time t_{i-1} which yielded the minimum transition penalties.

(5) Choose the optimal pitch-period trajectory. When the cumulative penalty has been calculated for the last time sample to be considered in the phrase, the pitch-period trajectory with the lowest cumulative penalty is chosen as the optimal trajectory.

(6) Quantize the optimal pitch-period trajectory. The resulting optimal pitch-period trajectory is then quantized to one of sixteen levels at

each time point on the trajectory. The quantized pitch period trajectory is then stored with the spectrum at that time point. The pitch period at each time sample is treated as an auxiliary measure and is used for verification exactly as the spectrum. The use of the pitch period in verification is described in Section VI.C.2. The quantization levels follow in Table 7.

TABLE 7. PITCH-PERIOD QUANTIZATION

Quantization Level	Pitch-Period Range (ms)	
0	$\tau \leq 6.0$	($\geq$ 167 Hz)
1	$6.0 < \tau \leq 6.45$	
2	$6.45 < \tau \leq 6.9$	
3	$6.9 < \tau \leq 7.35$	
4	$7.35 < \tau \leq 7.8$	
5	$7.8 < \tau \leq 8.25$	
6	$8.25 < \tau \leq 8.7$	
7	$8.7 < \tau \leq 9.15$	
8	$9.15 < \tau \leq 9.6$	
9	$9.6 < \tau \leq 10.05$	
10	$10.05 < \tau \leq 10.5$	
11	$10.5 < \tau \leq 10.95$	
12	$10.95 < \tau \leq 11.4$	
13	$11.4 < \tau \leq 11.85$	
14	$11.85 < \tau \leq 12.3$	
15	$\tau > 12.3$	(< 50 Hz)

SECTION VI
EXPERIMENT

A limited experiment was performed for purposes of comparing the performance of the compensation techniques over the original and degraded channels.

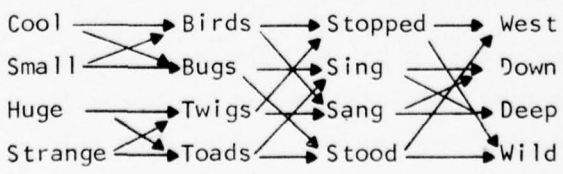
A. Data Set

The data set consisted of a subset of 16 speakers from the BISS Phase I data set collected at Texas Instruments. An enrollment session and four execution sessions were selected for each speaker. The enrollment session consisted of 20 phrases and each execution session contained four phrases. The phrases were prompted from a randomly selected set of 32 phrases containing four monosyllabic words. The words used in the BISS Phase I set are given in Table 8 with the allowable sequences shown by the arrows. Each true speaker was used as a casual impostor against the other 15 speakers.

B. Degraded Channel Condition

The analog data set of 16 speakers was processed through the RADC Digital Communications Experiment Facility (DICEF) to impose a 4A line characteristic to the data to simulate a telephone line. The DICEF facility is a digital filtering system that can introduce amplitude distortion and time delay characteristics are shown in Figures 6 and 7, respectively. The sampling frequency used in this experiment was 9200 samples per second. It was discovered after the data had been processed that it was not low-pass filtered at the half-sample frequency to prevent aliasing. This apparently caused some time registration problems which will be discussed in Section VII.A.

TABLE 8. WORD SET FOR VERIFICATION UTTERANCE CONSTRUCTION



White noise was added to the data to simulate a 20 dB signal-to-noise ratio. Less than one percent of the telephone lines have a signal-to-noise ratio as low as 20 dB, according to the Bell Laboratory Survey.[2] This was considered a good test for the noise resistance of the speaker verification strategy. The data processed through the 4A channel with white noise added to simulate a 20 dB signal-to-noise ratio will be referred to as the degraded channel.

C. Enrollment

The enrollment of the speakers means that an average set of reference scanning patterns and recognition patterns are obtained for each of the 16 words. There are 20 phrases of enrollment data selected so that each word is repeated five times. Therefore, the average pattern for each word is from five separate patterns.

1. Scanning Patterns

The scanning patterns are defined by first manually choosing two time-registration points for each word in the phrase. These registration points were chosen to lie just before the vowel and just after the vowel in a region where the spectrum has the most change. When the two registration points have been located, a scanning pattern is defined at each of the points, as described in Section II.B.1 and shown in Figure 1. Thus, each word has two scanning patterns associated with it.

For this experiment, the registration points are chosen manually for the first four phrases of the enrollment. Preliminary scanning patterns are then defined which are used to scan the remainder of the enrollment phrases to mark the registration points. As each phrase is marked in order, a new scanning pattern for each word in the phrase is defined and averaged with the previous patterns for that word.

2. Pitch

Once the registration points have been marked, the pitch period is extracted. Since the time registration for the words is chosen around the vowels, most of the region between the registration points should be voiced,

depending on the exact location of the registration points. The pitch period was then extracted only between the two registration points for each word.

3. Recognition Patterns

The recognition patterns for each word were defined by choosing six columns of the spectrum, plus pitch period, interpolated between the two marked time-registration points for that word. Again, five patterns are summed to obtain the reference recognition patterns for each word. These recognition patterns will be used to compute a spectral and a pitch-period error for verification purposes.

D. Execution

In the execution phase of the study, the input data is scanned by the reference scanning patterns belonging to the words in the phrase under consideration. As described earlier in Section II.B, an optimal decision strategy is used to choose a best set of eight registration points for each phrase. With the registration points thus defined, the pitch-period is extracted for the points between the registration points of each word. The pitch-period was then extended for five time samples before the first registration point of the word and five samples beyond the last registration point. This was to allow for some variance in the registration points when using scanning patterns defined on various channel conditions. The input data was then formatted into recognition patterns and compared with the reference patterns to form a spectral and pitch-period error for each phrase.

E. Design

The enrollment for the experiment was performed on data from the original channel. The execution was performed for data from both the original and the degraded channels. In addition, a test was run with registration on the degraded channel and execution on the original and the degraded channels to determine which channel condition should be used for enrollment. Several verification errors were then calculated for each execution phrase.

These were:

(1) Spectral error E_s

(2) Pitch-period error E_p

(3) Relative pitch-period error $E_{rp} = E_p - \bar{E}_p$, where $\bar{E}_p$ is the mean value of the pitch-period over the phrase.

In addition, two weighted sum errors were calculated:

(4) $E_s + W E_p$

(5) $E_s + W E_{rp}$, where W is a weighting constant.

The above error measures were used in several multiple phrase strategies. The definition of the one, two, three, and four phrase strategies is given in Table 9. In this table, the E_1 represents the error measure (this could be any one of the five measures given above; example, spectral error) for the first phrase of the four phrases of each execution session; E_2 is the same error measure for the second phrase of the execution session; E_3 is the error measure for the third phrase; and E_4 is the error measure for the fourth phrase.

TABLE 9. MULTIPLE PHRASE STRATEGIES

One phrase strategy: $\bar{E}_1 = E_1$; $\bar{E}_2 = E_2$; $\bar{E}_3 = E_3$; $\bar{E}_4 = E_4$

Two phrase strategy: $\bar{E}_1 = \frac{E_1 + E_2}{2}$; $\bar{E}_2 = \frac{E_2 + E_3}{2}$; $\bar{E}_3 = \frac{E_3 + E_4}{2}$;

$\bar{E}_4 = \frac{E_4 + E_1}{2}$

Three phrase strategy: $\bar{E}_1 = \frac{E_1 + E_2 + E_3}{3}$; $\bar{E}_2 = \frac{E_2 + E_3 + E_4}{3}$;

$\bar{E}_3 = \frac{E_3 + E_4 + E_1}{3}$; $\bar{E}_4 = \frac{E_4 + E_1 + E_2}{3}$

Four phrase strategy: $\bar{E}_1 = \bar{E}_2 = \bar{E}_3 = \bar{E}_4 = \frac{E_1 + E_2 + E_3 + E_4}{4}$

SECTION VII
RESULTS

The results of the limited-data test were very encouraging. Using a four phrase strategy with a weighted sum error of the spectral and relative pitch-period errors, a one percent true speaker rejection rate and one percent impostor acceptance rate were obtained. All of the true speaker data was able to be time-registered; however, some of these were misregistrations. The results will be discussed in more detail in the following sections.

A. Time Registration

All of the true-speaker data was able to be time-registered. However, it was determined that several of the phrases were registered at incorrect locations. Out of 256 phrases of true-speaker data, four phrases were grossly misregistered on at least one word and ten more phrases were judged to have minor misregistrations. An example of a minor misregistration is shown in Figure 8 for the word "sing" in the phrase "huge toads sing deep." Good registration points for the word "sing" would be at time points 179 and 192 which were obtained when the enrollment and execution were both on the degraded channel. When the enrollment was on the original channel and the execution on the degraded channel, the registration points were at 172 and 192. This was the fourth execution session of speaker No. 3. A possible explanation for the misregistration can be seen by looking at the enrollment of the word "sing" for speaker No. 3 over the original and degraded channels. Figure 9 shows the spectrum for the word "sing" on the original channel. The first registration point for the word was chosen at time sample 168. Notice the lack of energy just before this registration point. Comparing this to the registration over the degraded channel shown in Figure 10, it is apparent that considerable energy is present in front of the registration point at 175. This energy has been added by the 4A line and is thought to be due to the aliasing that is present. As mentioned previously, the DICEF facility did not low-pass filter the data before sampling, which caused aliasing to occur. It is not felt that the additional energy is due to the noise since

Execution Registration of "Sing"

Enroll - Original; Execute - Degraded (2)

Enroll - Degraded; Execute - Degraded (1)

Figure 8. Time Registration of the Word "Sing" on a Degraded Channel

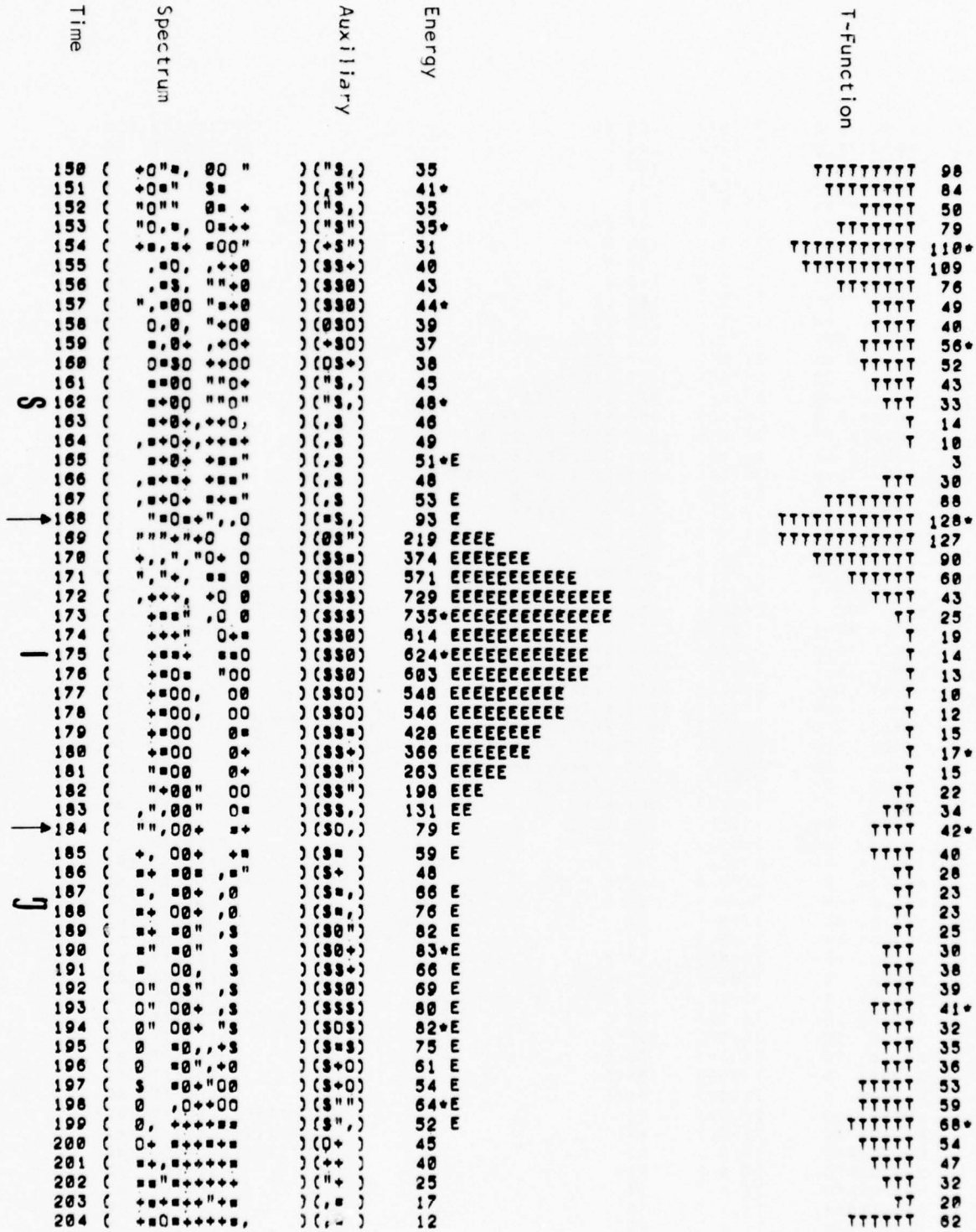


Figure 9. Enrollment Time Registration of "Sing" on Original Channel

Figure 10 Enrollment Time Registration of "Sing" on Degraded Channel

that seems to occur more or less uniformly throughout the word. It was thought at first that the aliasing would not cause a problem since the verification is done on the vowel that is relatively low frequency. However, the registration points are placed at rapidly changing points of the spectrum and, in this case and several others, the registration is between either a fricative or stop consonant and the vowel. The high frequency content of the fricative or stop causes the problem.

Also, in analyzing the results of the registration, it became apparent that the pitch-period could be used as a method of eliminating gross misregistrations. To see this, consider that the registration points are chosen about the vowels which are voiced. These registration points are chosen just before and just after the vowel so the pitch track may not be stable at the beginning and the end; however, most of the interval between the registration points should be voiced. It has been found that if little voicing is present in the interval, registration points are badly placed and the pitch-period error is usually very large. Therefore, by requiring that voicing occur over some percentage of the interval, such as 60%, it can be assumed that the registration is probably correct. The voicing decision would be very easy to calculate at the time the pitch is being extracted.

B. Verification

The first result of interest is the determination of the channel condition to use for enrollment. Table 10 gives the results for the spectral error and the pitch-period error of enrolling on the original channel and executing on both the original and degraded channels. Likewise, the results of enrolling on the degraded channel with execution on both the original and degraded channels are given. The results are given in the form of the equal error rate, which is the rate at which the true speaker rejection and impostor acceptance are equal. As can be seen from the table, if execution is always to be over the same line as the enrollment, then that line should be used for enrollment. However, if any cross-channel conditions are to be encountered (i.e., enrollment on one channel condition and execution over a different channel), the results indicate the best overall performance is obtained by enrolling on the original channel.

Hence, the enrollments on the remainder of the experiment were performed using the original channel conditions. The results for the remainder of the experiment also include the multiple phrase strategies discussed in Section VI.E.

TABLE 10. EQUAL ERROR RATES FOR ENROLLMENT ON ORIGINAL AND DEGRADED CHANNELS

Attribute		Enroll	Execute	Equal Error Rate
Spectral Error	$E_{spectral}$	Original	Original	5.0
		Original	Degraded	6.4
		Degraded	Original	6.8
		Degraded	Degraded	6.1
Pitch-Period Error	E_{pitch}	Original	Original	25.5
		Original	Degraded	26.2
		Degraded	Original	26.7
		Degraded	Degraded	25.2

The equal error rate results for the spectral error are shown in Table 11. The 14-channel results given at the top of the table are taken from the Speaker Verification II Report.[3] The 10-channel results were obtained in the current experiment. The differences in the results are not entirely due to the number of channels of spectral data used. The 14-channel results were obtained from an experiment using the entire BISS Phase I data set and processing developed for the BISS system. These are described in the Speaker Verification II Report.[3] The 10-channel results were obtained using a subset of the BISS Phase I data set and processing described in Sections VI.A and V, respectively, of this report.

The decision function distribution using the spectral error in a one-phrase strategy is given in Figure 11. The figure compares the distributions obtained from executing over the original channel and the degraded channel. Figure 12 shows the decision function distributions of the spectral error for a four-phrase strategy. Again, the comparison is of the distributions

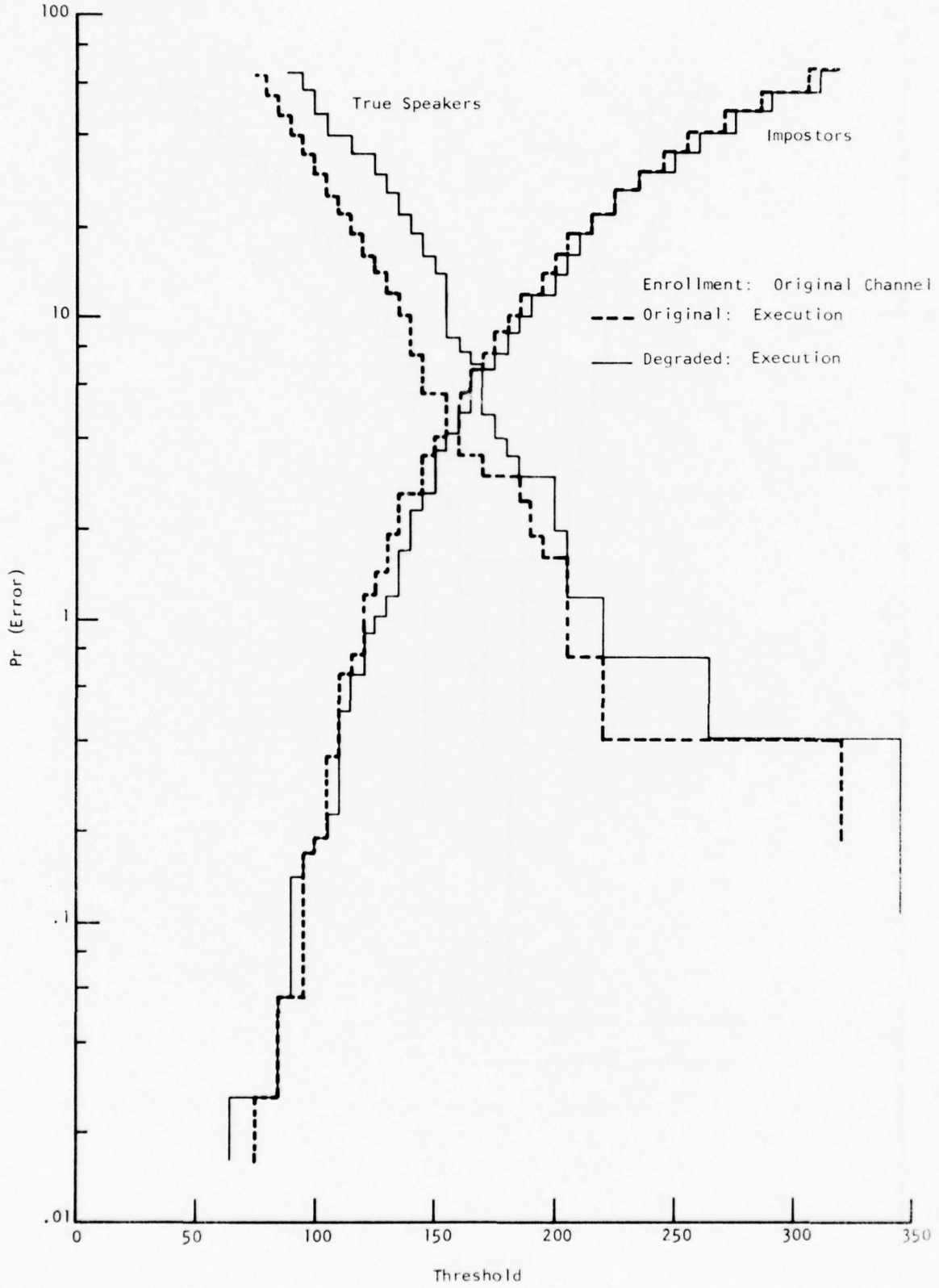


Figure 11. Decision Function Distributions Using Spectral Error in One Phrase Strategy

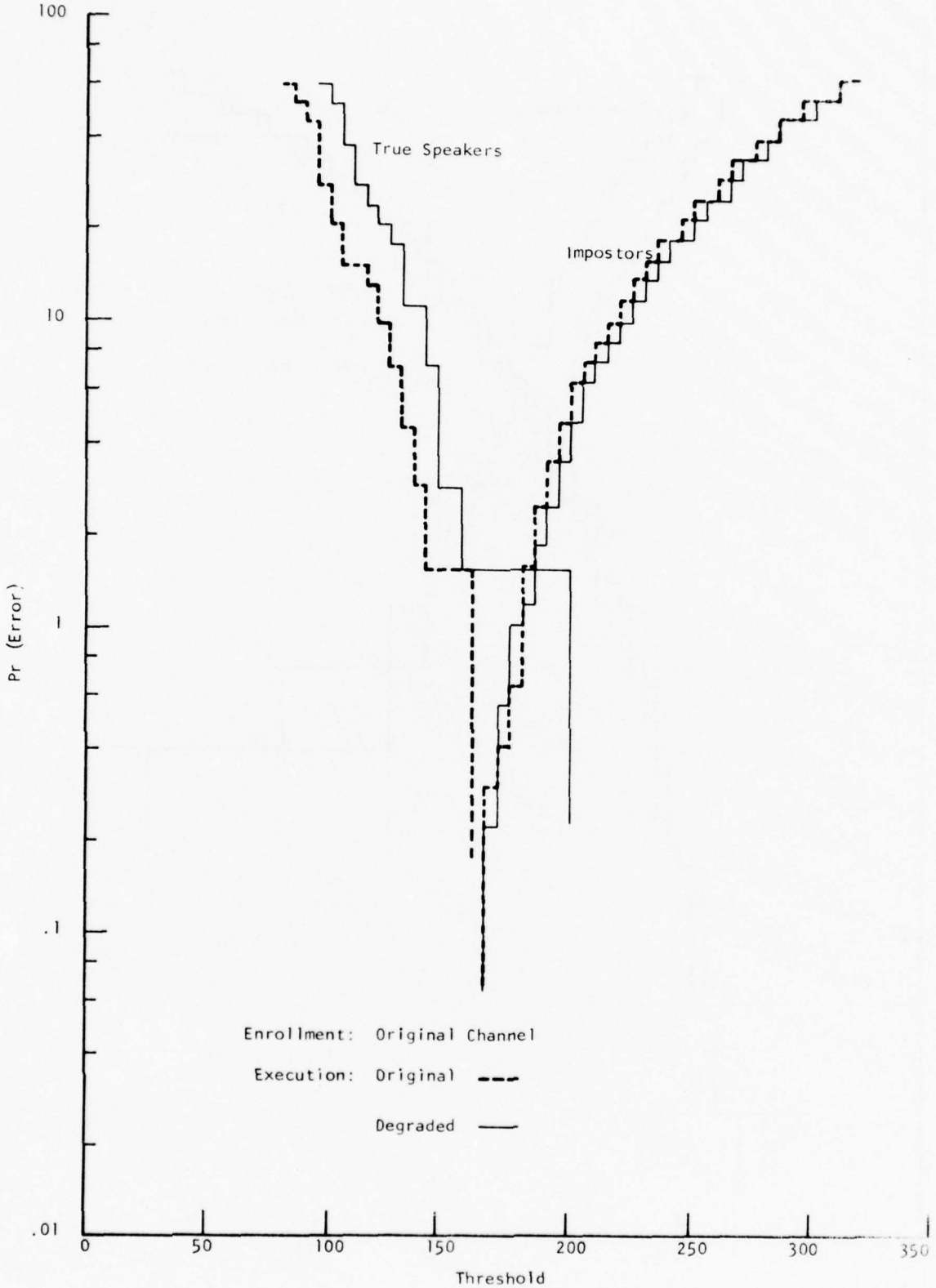


Figure 12. Decision Function Distributions Using Spectral Error in Four Phrase Strategy

obtained from executing on the original channel and on the degraded channel. Note that the BISS specifications of one percent true speaker rejection and two percent impostor acceptance are not met with the spectral error when executing over a degraded channel.

TABLE 11. EQUAL ERROR RATE FOR MULTIPHRASE STRATEGIES USING THE SPECTRAL ERROR

Attribute	Enrollment	Execution	# Phrases	Equal Error Rate
Spectral Error (14-channel)	Original	Original	1	4.2
			2	1.6
			3	0.9
			4	0.5
Spectral Error (10-channel)	Original	Original	1	5.0
			2	1.5
			3	1.0
			4	0.3
Spectral Error (10-channel)	Original	Degraded	1	6.4
			2	2.7
			3	1.6
			4	1.6

The equal error rates for the pitch-period error and the relative pitch-period error are given in Table 12. There are several items of particular interest in this table. First, the performance for the pitch-period error on the original and degraded channels is approximately equivalent, as was expected. However, this doesn't seem to hold true for the relative pitch-period error. This was not due to any failing of the relative pitch-period error, but because three of the sequences used were grossly misregistered. This caused a very large error to occur in the relative pitch-period, but not as severe an error in the pitch-period. The problems of misregistration were covered in the previous section.

TABLE 12. EQUAL ERROR RATE FOR MULTIPHRASE STRATEGIES USING THE PITCH-PERIOD AND RELATIVE PITCH-PERIOD ERRORS

Attribute	Enrollment	Execution	# Phrases	Equal Error Rate
Pitch-Period Error	Original	Original	1	25.5
			2	20.8
			3	19.2
			4	17.2
Pitch-Period Error	Original	Degraded	1	26.2
			2	22.0
			3	19.1
			4	18.4
Relative Pitch-Period Error	Original	Original	1	18.9
			2	13.4
			3	9.2
			4	6.1
Relative Pitch-Period Error	Original	Degraded	1	19.3
			2	14.9
			3	15.5
			4	13.8

A second item of interest in the table is that the relative pitch-period error gave better overall performance than did the pitch-period error. This is because the relative pitch-period error approximates the performance of the pitch-period slope. Doddington[4] found that better performance is obtained with the pitch-period slope than with the pitch-period.

A third item of interest also relates to the results of Doddington.[4] In the experiments using pitch-period slope, he achieved an equal error rate of six percent on a one-phrase strategy as opposed to the 19% obtained here. The various reasons for this are discussed below.

(a) In his experiments, Doddington adapted the speaker's pitch-period slope over as many as ten sessions with two-day lapses between

sessions. In the experiment conducted here, the enrollment on the pitch-period was accomplished in one session. The pitch-period can vary from enrollment to execution as shown in Figure 13, where a plot of the average pitch-period for several words is plotted versus the occurrence of the word. The first five occurrences were during enrollment and the pitch-periods are relatively stable for the two speakers. However, in the next four occurrences of the word, which were during the execution sessions, a considerable variation in the pitch period is observed. This variation was probably due to "mike fright" and would dissipate as the speaker became familiar with the system. This variation problem can be overcome with an adaptation of the pitch-period over several sessions.

(b) The phrase used in Doddington's experiment was a single phrase, "We were away a year ago"; whereas, the phrases used here are a random assortment of four monosyllabic words. The pitch-period obtained for any one word in the phrase might change depending on the two words immediately preceding and following. The enrollment sessions did not allow all of the transitions of the words to occur. This problem can also be overcome with adaptation of the speaker's pitch period.

(c) The phrases used in the current experiment were prompted. This could have an adverse effect on performance in that some of the speakers might change their manner of speaking to accommodate the prompting. Adapting the speaker's pitch-period would also tend to alleviate this problem.

The decision function distributions using the pitch-period error in a four-phrase strategy are plotted in Figure 14, and the decision function distributions using the relative pitch-period error in one and four-phrase strategies are given in Figure 15 and 16, respectively. In all three figures, the distributions are compared for execution over the original and degraded channel conditions.

Two weighted sum errors were then calculated for various values of a weighting constant W. These weighted sum errors were the spectral error

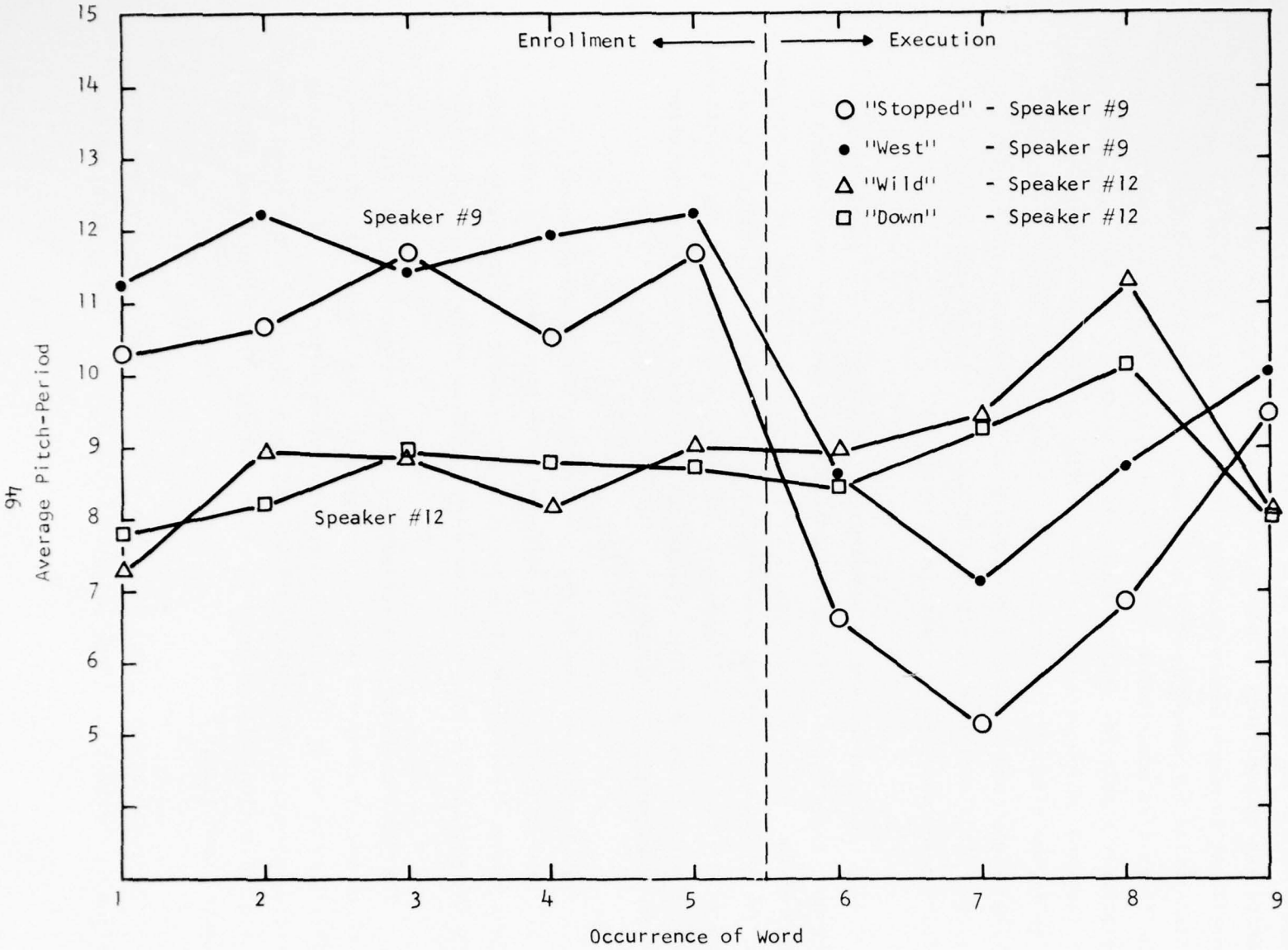


Figure 13. Variations in Pitch-Period Versus Occurrence

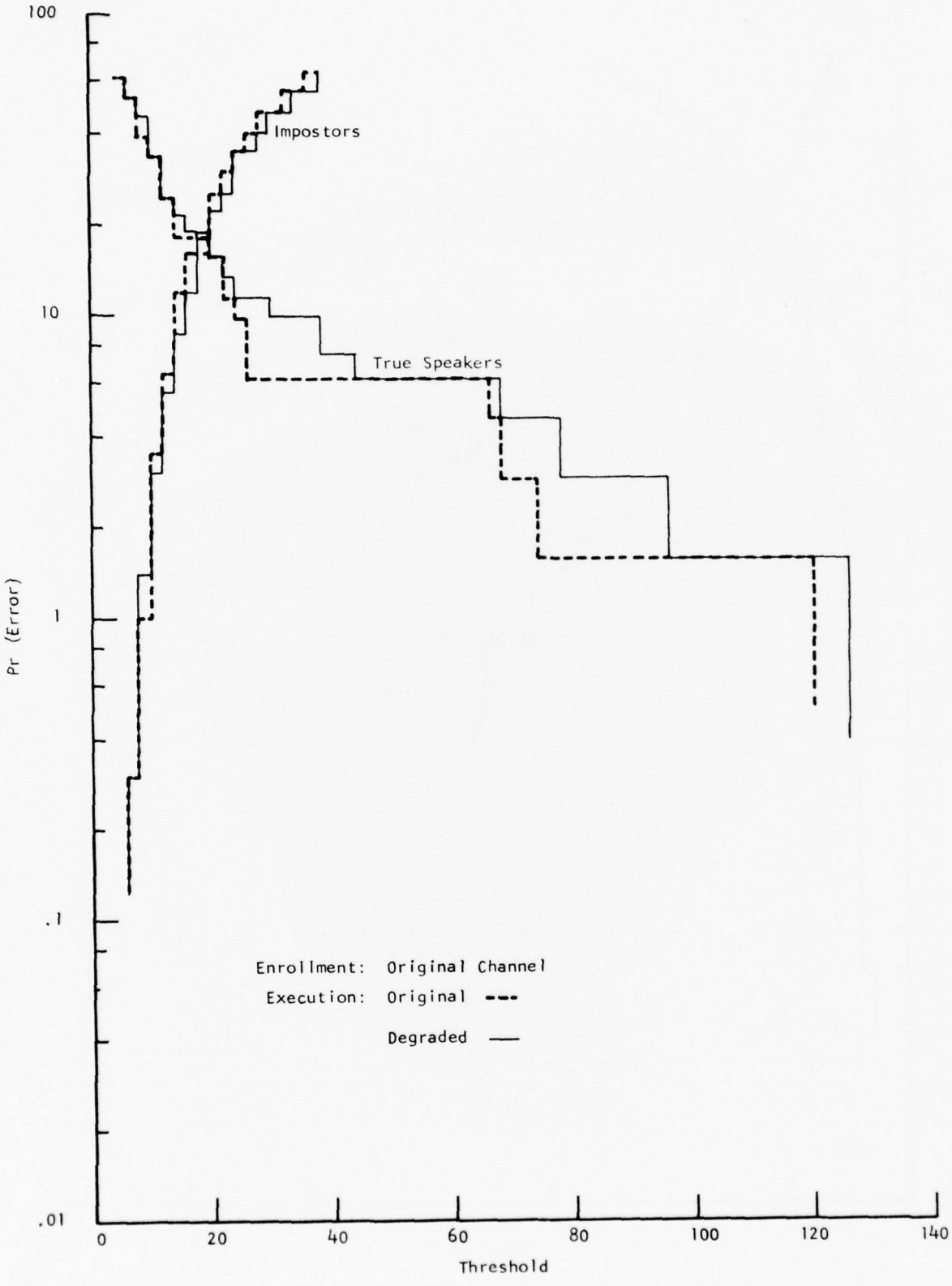


Figure 14. Decision Function Distributions Using Pitch-Period Error with Four Phrase Strategy

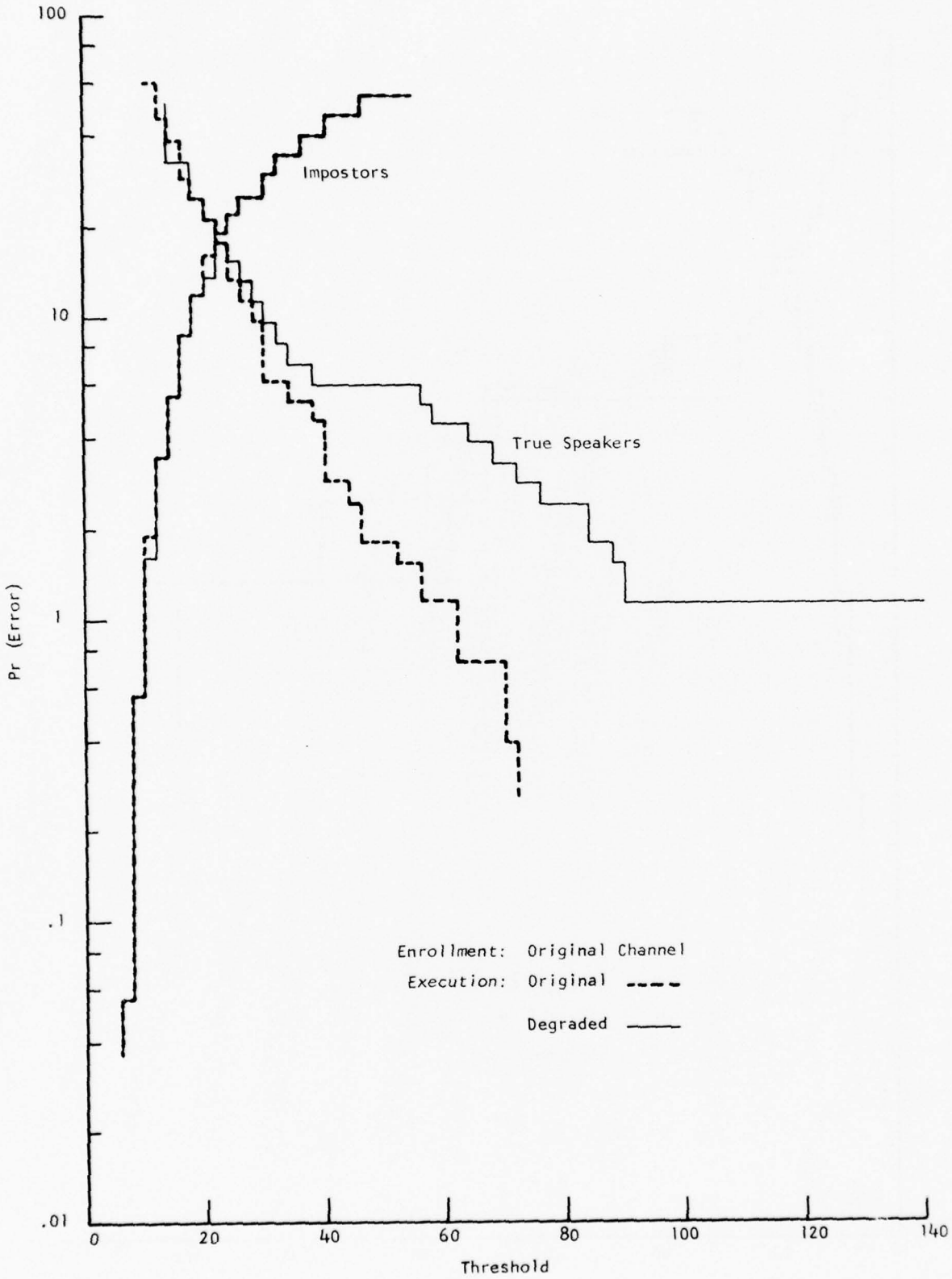


Figure 15. Decision Function Distribution Using Relative Pitch-Period Error with One Phrase Strategy

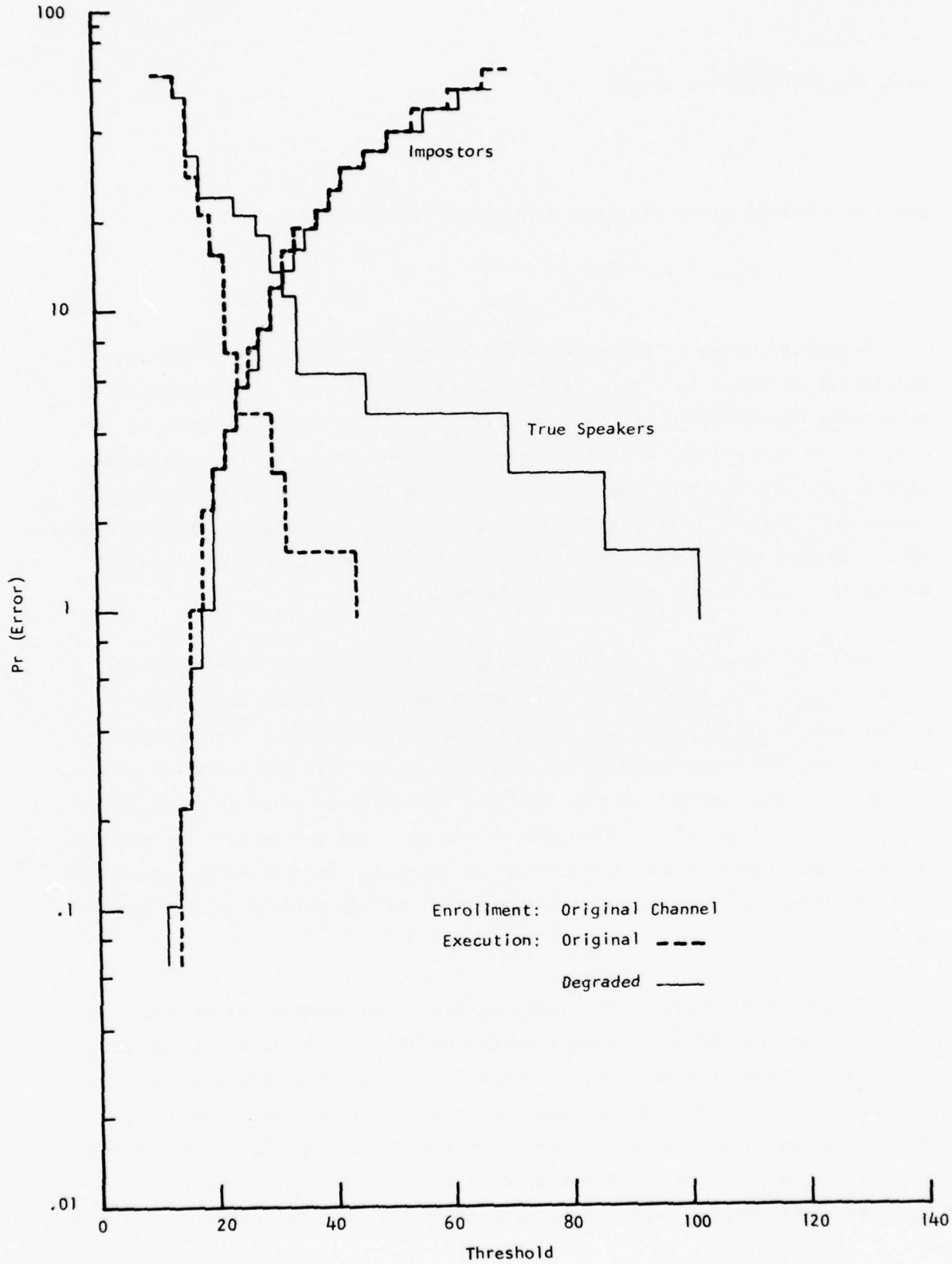


Figure 16. *Decision* Function Distributions Using Relative Pitch-Period with Four Phrase Strategy

plus the pitch-period error

$$E_{sp} = E_s + W E_p \quad ,$$

and the spectral error plus the relative pitch-period error

$$E_{srp} = E_s + W E_{rp} \quad .$$

A typical curve of the equal error rate, EER, versus the weighting constant W is shown in Figure 17 for the particular case of the spectral error plus the relative pitch-period error E_{srp} with the enrollment on the original channel, execution on a degraded channel, and various multiphrase strategies. Notice that the optimum weighting W_{opt} which is the minimum of the curves, varies considerably from one strategy to another. This randomness of the optimum weighting constant for the various strategies is probably due to the limited data set used in the experiment.

Table 13 shows the equal error rate for the two weighted sum errors E_{sp} and E_{srp}. The equal error rates shown were obtained by using the optimal weighting constant W_{opt} for each of the multiphrase strategies. As can be seen, the combination of the spectral error with the relative pitch-period error gave better results than did the spectral error with the pitch-period error. Also, the performance of the weighted sum errors was worse on the degraded channel than on the original channel. Part of this degradation in performance was due to the misregistration of the phrases as mentioned earlier.

The decision function distributions using the spectral error plus the relative pitch-period error with a weighting of $W = 0.95$ in a four-phrase strategy is shown in Figure 18. A comparison is shown of the distributions for execution with the degraded and the orignal channel conditions. By choosing a decision threshold D_t which is such that $215 < D_T < 220$, it can be seen that a performance of better than a one percent true speaker rejection and a one percent impostor acceptance is obtained for execution on both

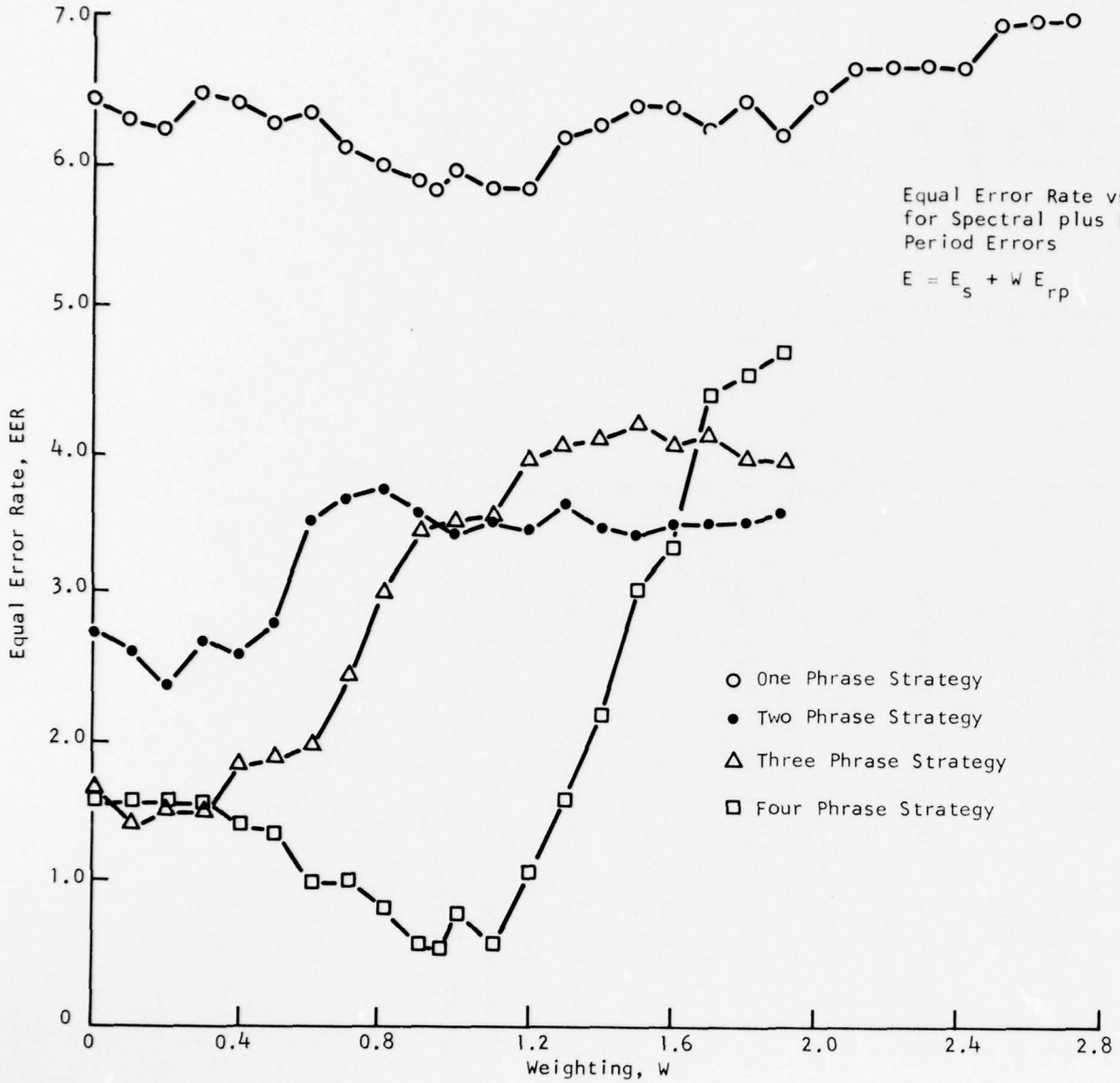


Figure 17. Equal Error Rate Versus Weighting for Spectral Plus Relative Pitch-Period Error

the original and degraded channels. Thus it has been shown that for the limited experiment conducted here, the BISS specifications of one percent true speaker rejection and two percent impostor acceptance were met.

TABLE 13. EQUAL ERROR RATE FOR MULTIPHRASE STRATEGIES USING WEIGHTED SUM ERRORS

Attribute	Enrollment	Execution	# Phrases	Equal Error Rate	Weighting (W)
Spectral + W	Original	Original	1	4.7	0.35
(Pitch-Period			2	1.3	0.55
Error): E_{sp}			3	1.0	0.0
			4	0.2	0.05
Spectral + W	Original	Degraded	1	6.0	0.65
(Pitch-Period			2	2.4	0.4
Error): E_{sp}			3	1.5	0.1
			4	1.6	0.65
Spectral + W	Original	Original	1	3.9	0.5
(Relative			2	1.0	0.8
Pitch-Period			3	0.5	1.8
Error): E_{srp}			4	0.0	0.25
Spectral + W	Original	Degraded	1	5.8	0.95
(Relative			2	2.4	0.2
Pitch-Period			3	1.4	0.1
Error): E_{srp}			4	0.6	0.95

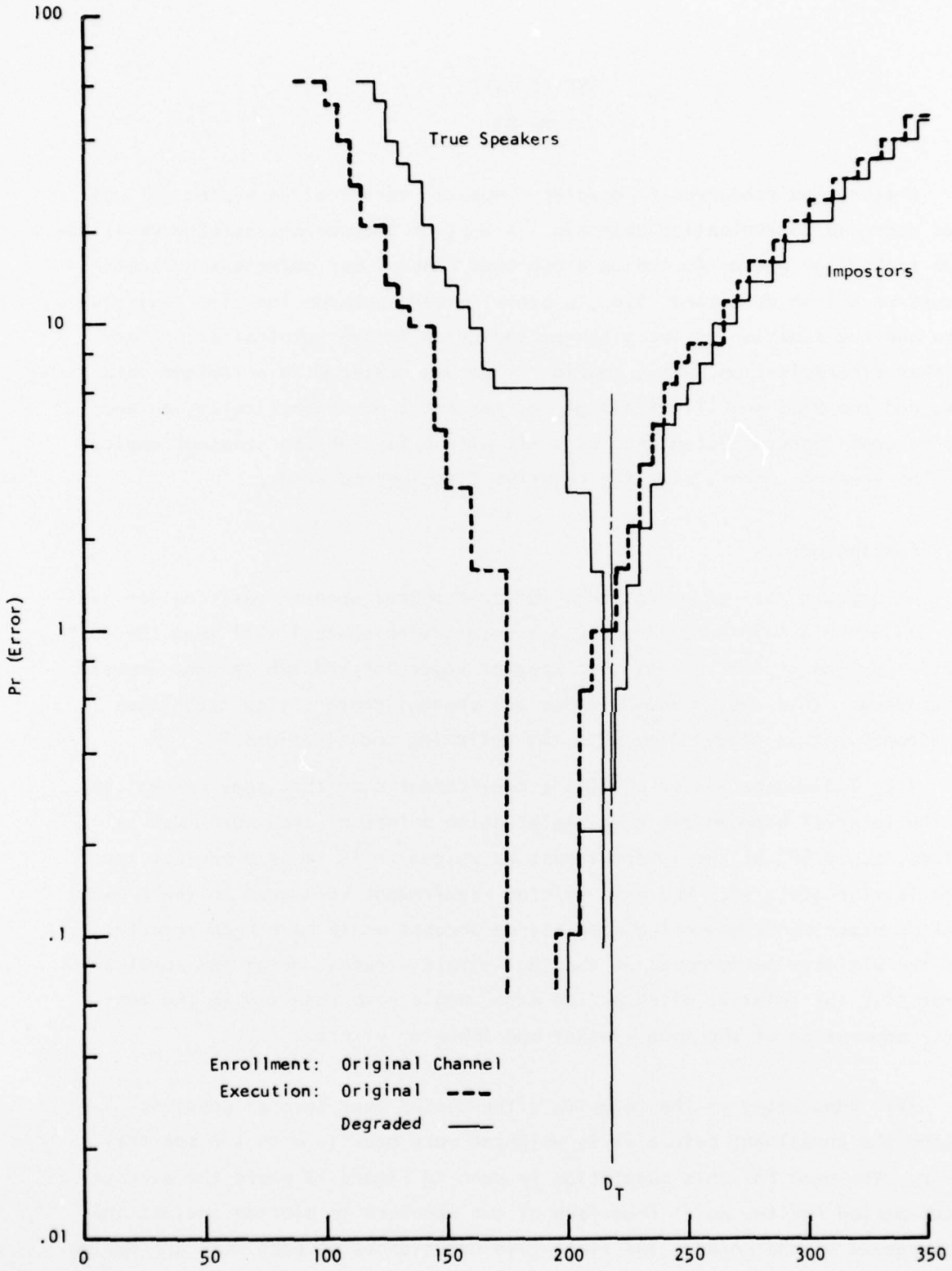


Figure 18. Decision Function Distributions Using Spectral Plus Relative Pitch-Period Error with W = 0.95 in Four Phrase Strategy

SECTION VIII
SUMMARY

A study was conducted to develop a speaker verification system for use over degraded communication channels. A test of the current speaker verification technology was performed on a degraded channel and compensation techniques were then developed, i.e., a band-limited spectrum for time registration and the addition of the pitch-period error to the spectral error for speaker discrimination. This configuration was tested with a limited data set, and the BISS specifications of one percent true speaker rejection and two percent impostor acceptance were met with a four-phrase strategy employing the spectral error, plus the relative pitch-period error.

A. Conclusions

It appears that an operational remote terminal speaker verification system utilizing a telephone line for a communication channel will meet the BISS specifications of one percent true speaker rejection and two percent impostor acceptance. This system would employ the channel compensation techniques developed in this study along with the following modifications.

(1) Implementation of a voicing requirement such that some percentage of the interval between the time registration points of each word must be voiced, e.g., 60% of the interval must be voiced or it is assumed that the word is misregistered. Had this voicing requirement been used in the present limited experiment, several misregistered phrases would have been rejected and the ultimate performance of the four-phrase strategy using the spectral error plus the relative pitch-period error would have resulted in the complete separation of the true speaker and impostor errors.

(2) Adaptation of the relative pitch-period over several sessions beyond the enrollment before it is weighted very heavily with the spectral error. The need for this adaptation is seen in Figure 13 where the average pitch-period for two words from each of two speakers is plotted against the occurrences of the words. The first five occurrences of each word are during

the enrollment session and a consistent pitch-period is observed for each word. However, during the next four occurrences, which are the execution sessions, a wide variance of pitch periods is observed. Averaging the pitch-period over the execution sessions, as well as over the enrollment, would result in a much better reference pitch-period for each word. This would then lead to much improved performance of the weighted sum error for speaker discrimination.

(3) Normalization of the spectral error by the noise energy from regions of the spectrum in which the speaker is silent. This would tend to stabilize the spectral error from the original to the degraded channel and would make it easier to pick a decision threshold.

It can be seen from Figure 18 that a decision threshold D_T which is such that $215 < D_T < 220$, results in one percent true speaker rejection and one percent impostor acceptance for execution over both original and degraded channel conditions. This performance was achieved using a four-phrase strategy with the spectral error plus the relative pitch-period error. The problem of choosing a decision threshold was not addressed to any great extent, however, because of the limited data set used in the experiment. To set the decision thresholds for an operational system requires a larger experiment.

B. Telephone Experiment

To obtain a preliminary evaluation over an actual telephone, Dan Daniel of Texas Instruments (a participant in the original BISS Phase I experiment) was asked to enroll over a telephone in the laboratory and to use four different office telephones in the laboratory for execution sessions. Eight phrases of execution were performed in each of the four offices by dialing an inside line (Centrex) and also by dialing an outside line (dial "9"). In one of the offices, he was also asked to call the Texas Instruments plant at Austin, Texas, and have them patch him back to the data phone in the lab. He also performed an execution session in the sound booth as a reference, since his original data was collected in the sound booth. The results of the

eight phrases of execution were averaged and the results are given in Table 14 showing the average spectral and pitch-period errors. The variations in the performance between the inside lines and the outside lines is due to the different switching equipment being used for the two lines. The variations observed from office to office using only the inside line or the outside line, however, were due to the variations in the telephone headset microphones. This is true because all of the offices are connected to the same telephone line through a "rotary" system. Thus, if an inside line is called from office 1 and 2, the same line and switching network are being used, but the headset is different.

A very encouraging aspect of this limited experiment was that all of the execution phrases were correctly time-registered.

C. Recommendations for Future Work

To verify the performance of the remote terminal speaker verification system developed in this study, a larger simulated telephone experiment should be conducted using the 4A line with a 20 dB signal-to-noise ratio. The system to be tested would include the modifications discussed in Section VIII.A, along with the channel compensation techniques developed in this study. Before this experiment can be conducted, however, the following two preliminary tasks must be completed.

(1) A method of automatically enrolling the speakers on the system must be developed.

(2) A real time pitch tracker must be implemented. An array processor will be installed on the laboratory system in midyear 1977 which will allow this task to be accomplished.

In addition to an evaluation of the performance of the speaker verification system, the simulated telephone experiment should provide a set of decision thresholds and a verification strategy.

Upon the completion and evaluation of the simulated telephone experiment, another experiment should be conducted over actual telephone lines. This would be a two-fold experiment. The first phase would consist of a limited scale experiment to study the headset microphone problem. This experiment would be conducted using various telephones with both the existing microphones and with these microphones replaced by high quality microphones. Depending on the outcome of this limited experiment, a large scale experiment would be conducted over the telephone channels with either the existing or high quality microphones.

TABLE 14. DAN DANIEL TELEPHONE EXPERIMENT

Enrollment	Execution (8 Phrases)	Avg. Spectral Error	Avg. Pitch Error	Avg. Spectral + Pitch Error
Orig. (1974) (Sound Booth)	Sound Booth	79	2.5	81.5
	Inside - Office 1	85.4	3.8	89.2
	Dial 9 - Office 1	109	4.4	113.4
	Inside - Office 2	163	5.0	168.0
	Dial 9 - Office 2	110	4.6	114.6
	Inside - Office 3	234	4.1	238.1
	Dial 9 - Office 3	129	4.1	133.1
	Inside - Office 4	183	4.1	187.1
	Dial 9 - Office 4	142	5.0	147.0
	Austin - Office 4	124	2.5	126.5
Dial 9 - Office 1	Sound Booth	119.4	5.9	125.3
	Inside - Office 1	72.8	4.0	76.8
	Dial 9 - Office 1	52.3	2.4	54.7
	Inside - Office 2	197	2.6	199.6
	Dial 9 - Office 2	107.1	2.4	109.5
	Inside - Office 3	269.0	3.5	272.5
	Dial 9 - Office 3	111.3	2.9	114.2
	Inside - Office 4	215.4	2.0	217.4
	Dial 9 - Office 4	132.4	2.5	134.9
	Austin - Office 4	133.5	4.5	138.0

REFERENCES

1. G. R. Doddington, R. E. Helms, B. M. Hydrick, "Speaker Verification III," Final Report for Rome Air Development Center, 6 February 1976.

2. F. P. Duffy and T. W. Thatcher, Jr., "1969-1970 Connection Survey: Analog Transmission Performance on the Switched Telecommunications Network," Bell Syst. Tech. J. 50, April (1971).

3. G. R. Doddington and B. M. Hydrick, "Speaker Verification II," Final Report for Rome Air Development Center, 15 September 1975.

4. G. R. Doddington, "A Method of Speaker Verification," Ph.D. dissertation, University of Wisconsin, 1970.

5. B. S. Atal, "Automatic Speaker Recognition Based on Pitch Contours," J. Acoust. Soc. Am. 52, 1687 (1972).

APPENDIX

COMPARISON OF BISS AND REMOTE TERMINAL SPEAKER VERIFICATION SYSTEM PROCESSING

The differences between the BISS automatic speaker verification system and the remote terminal speaker verification system are presented in this appendix.

A. Preprocessing

1. Filter Bank Definition

	BISS (Digital)			Remote Terminal (Analog)	
Channel No.	Center Freq. (Hz)	Bandwidth (Hz)	Channel No.	Center Freq. (Hz)	Bandwidth (Hz)
1	355	220	1	350	300
2	530	220	2	450	300
3	705	220	3	555	310
4	880	220	4	670	340
5	1055	220	5	790	380
6	1230	220	6	940	400
7	1405	220	7	1120	400
8	1580	220	8	1320	400
9	1755	220	9	1550	440
10	1930	220	10	1810	400
11	2105	220	11	2100	400
12	2280	220	12	2420	400
13	2455	220	13	2800	400
14	2630	220	14	3200	400
15	2805	220	15	3800	800
16	2980	220	16	5000	1600

2. Regression

BISS

$$(\bar{A}_j)_R = \bar{A}_j - \sum_{k=1}^{2} C_{kj} \bar{P}_k$$

Remote Terminal

$$(\bar{A}_j)_R = \bar{A}_j - \sum_{k=0}^{2} C_{jk} \bar{F}_k$$

BISS

where $\{P_k\}$ are discretized, legendre polynomials.

Remote Terminal

where

$$\bar{F}_k = \begin{Bmatrix} f_{1k} \\ . \\ . \\ . \\ f_{14,k} \end{Bmatrix} \quad k = 0, 1, 2$$

$$f_{i0} = 1$$

$$f_{i1} = -\sin\left[\frac{(i - 1/2)\,\pi}{14}\right]$$

$$f_{i2} = -\cos\left[\frac{(i - 1/2)\pi}{14}\right]$$

$$i = 1, \ldots, 14 \ .$$

3. Normalization

BISS

$$(\bar{A}_j)_N = \frac{(\bar{A}_j)_R}{\mu_j}$$

where $\mu_j = \sum_{i=1}^{14} (a_{ij})_R$ is the average value.

Remote Terminal

$$(\bar{A}_j)_N = \frac{(\bar{A}_j)_R}{\sigma_j^*}$$

where $\sigma_j^* = \sigma_{post_j} + \frac{1}{8} \max(\sigma_{post_n})$:

$$n = j - 8, \ldots, j + 8$$

$$\sigma_{post_j} = \frac{1}{11}\left(\sum_{m=1}^{14} a_{mj}^2 - \sum_{k=0}^{2} c_{jk}^2\right)$$

is the standard deviation.

By normalizing by the standard deviation rather than the mean, the resistance to noise has been increased by approximately 10 dB.[3]

4. Quantization

The method of quantization for the BISS and remote terminal systems is the same. The regressed and normalized vectors are quantized to one of

eight levels with the quantization threshold being computed to provide a uniform first-order distribution, i.e.:

$$\Pr[(a_{ij})_N \in (\phi_{i,q}, \phi_{i,q+1})] = \frac{1}{8} \text{ for all } q.$$

B. Processing

1. Scanning Pattern Definition

BISS	Remote Terminal
The scanning patterns are formed from the spectrum; the scanning pattern at time, t_j, consisting of six vectors of spectral data: $(A_{j-5}, A_{j-3}, A_{j-1}, A_{j+1}, A_{j+3}, A_{j+5})$.	The scanning patterns are formed from the scanning spectrum: the scanning pattern formed at time, t_j, consisting of (a) the average of spectral data at times, t_{j-1} and t_{j-2}; (b) the average of spectral data at times, t_{j+1} and t_{j+2}; and (c) the difference (b) - (a).

2. Verification Pattern Definition

BISS	Remote Terminal
The verification pattern is just the scanning pattern. The scanning error is used as the verification error. This can be thought of as a zeroeth order time warping to account for time alignment.	The verification pattern is six columns of spectral data which is interpolated between two time registration points. This is equivalent to a first order time warping in that it helps account for changes in the length of words in addition to time alignment.

3. Decision Function

BISS

$$d_N = \frac{\sum_{k=1}^{N} \sum_{i=1}^{4} E_{ik}}{\min\left[\max\left(\sum_{k=1}^{N} \sum_{i=1}^{4} \hat{E}_{ik}, 4\hat{NE}_{min}\right), 4\hat{NE}_{max}\right]}$$

where E_{ik} is the scanning error, $\hat{E}_{ik}$ is the expected scanning error, $\hat{E}_{min} = 100$, and $\hat{E}_{max} = 140$.

Remote Terminal

$d_N = E$ where E is the squared error between the reference and input verification patterns.

A decision strategy was not developed for the remote terminal study due to the limited experiment which was conducted.

METRIC SYSTEM

BASE UNITS:

Quantity	Unit	SI Symbol	Formula
length	metre	m	...
mass	kilogram	kg	...
time	second	s	...
electric current	ampere	A	...
thermodynamic temperature	kelvin	K	...
amount of substance	mole	mol	...
luminous intensity	candela	cd	...

SUPPLEMENTARY UNITS:

plane angle	radian	rad	...
solid angle	steradian	sr	...

DERIVED UNITS:

Acceleration	metre per second squared	...	m/s
activity (of a radioactive source)	disintegration per second	...	(disintegration)/s
angular acceleration	radian per second squared	...	rad/s
angular velocity	radian per second	...	rad/s
area	square metre	...	m
density	kilogram per cubic metre	...	kg/m
electric capacitance	farad	F	A·s/V
electrical conductance	siemens	S	A/V
electric field strength	volt per metre	...	V/m
electric inductance	henry	H	V·s/A
electric potential difference	volt	V	W/A
electric resistance	ohm		V/A
electromotive force	volt	V	W/A
energy	joule	J	N·m
entropy	joule per kelvin	...	J/K
force	newton	N	kg·m/s
frequency	hertz	Hz	(cycle)/s
illuminance	lux	lx	lm/m
luminance	candela per square metre	...	cd/m
luminous flux	lumen	lm	cd·sr
magnetic field strength	ampere per metre	...	A/m
magnetic flux	weber	Wb	V·s
magnetic flux density	tesla	T	Wb/m
magnetomotive force	ampere	A	...
power	watt	W	J/s
pressure	pascal	Pa	N/m
quantity of electricity	coulomb	C	A·s
quantity of heat	joule	J	N·m
radiant intensity	watt per steradian	...	W/sr
specific heat	joule per kilogram-kelvin	...	J/kg·K
stress	pascal	Pa	N/m
thermal conductivity	watt per metre-kelvin	...	W/m·K
velocity	metre per second	...	m/s
viscosity, dynamic	pascal-second	...	Pa·s
viscosity, kinematic	square metre per second	...	m/s
voltage	volt	V	W/A
volume	cubic metre	...	m
wavenumber	reciprocal metre	...	(wave)/m
work	joule	J	N·m

SI PREFIXES:

Multiplication Factors	Prefix	SI Symbol
1 000 000 000 000 = 10^{12}	tera	T
1 000 000 000 = 10^{9}	giga	G
1 000 000 = 10^{6}	mega	M
1 000 = 10^{3}	kilo	k
100 = 10^{2}	hecto*	h
10 = 10^{1}	deka*	da
0.1 = 10^{-1}	deci*	d
0.01 = 10^{-2}	centi*	c
0.001 = 10^{-3}	milli	m
0.000 001 = 10^{-6}	micro	μ
0.000 000 001 = 10^{-9}	nano	n
0.000 000 000 001 = 10^{-12}	pico	p
0.000 000 000 000 001 = 10^{-15}	femto	f
0.000 000 000 000 000 001 = 10^{-18}	atto	a

* To be avoided where possible.

MISSION
of
Rome Air Development Center

RADC plans and conducts research, exploratory and advanced development programs in command, control, and communications (C^3) activities, and in the C^3 areas of information sciences and intelligence. The principal technical mission areas are communications, electromagnetic guidance and control, surveillance of ground and aerospace objects, intelligence data collection and handling, information system technology, ionospheric propagation, solid state sciences, microwave physics and electronic reliability, maintainability and compatibility.